

SEQUENCE DETERMINATION OF NUCLEIC ACIDS USING ARRAYS WITH MICROSPHERES

This application is a continuation-in-part application of U.S.S.N.s 60/130,089, filed April 20, 1999;
60/135,051, filed May 20, 1999; 60/135,053, filed May 20, 1999, and 60/135,123, filed May 20, 1999.

FIELD OF THE INVENTION

The present invention is directed to methods and compositions for the use of microsphere arrays to
determine the sequence of nucleic acids, particularly alterations such as nucleotide substitutions
(mismatches) and single nucleotide polymorphisms (SNPs).

BACKGROUND OF THE INVENTION

The detection of specific nucleic acids is an important tool for diagnostic medicine and molecular
biology research. Gene probe assays currently play roles in identifying infectious organisms such as
bacteria and viruses, in probing the expression of normal and mutant genes and identifying mutant
genes such as oncogenes, in typing tissue for compatibility preceding tissue transplantation, in
matching tissue or blood samples for forensic medicine, and for exploring homology among genes
from different species.

Ideally, a gene probe assay should be sensitive, specific and easily automatable (for a review, see
Nickerson, Current Opinion in Biotechnology 4:48-51 (1993)). The requirement for sensitivity (i.e. low
detection limits) has been greatly alleviated by the development of the polymerase chain reaction
(PCR) and other amplification technologies which allow researchers to amplify exponentially a specific
nucleic acid sequence before analysis (for a review, see Abramson et al., Current Opinion in
Biotechnology, 4:41-47 (1993)).

Specificity, in contrast, remains a problem in many currently available gene probe assays. The extent

of molecular complementarity between probe and target defines the specificity of the interaction. Variations in the concentrations of probes, of targets and of salts in the hybridization medium, in the reaction temperature, and in the length of the probe may alter or influence the specificity of the probe/target interaction.

5

It may be possible under some circumstances to distinguish targets with perfect complementarity from targets with mismatches, although this is generally very difficult using traditional technology, since small variations in the reaction conditions will alter the hybridization. New experimental techniques for mismatch detection with standard probes include DNA ligation assays where single point mismatches prevent ligation and probe digestion assays in which mismatches create sites for probe cleavage.

10

Recent focus has been on the analysis of the relationship between genetic variation and phenotype by making use of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of single nucleotide polymorphisms (SNPs), which occur at an average frequency of more than 1 per kilobase in human genomic DNA. Some SNPs, particularly those in and around coding sequences, are likely to be the direct cause of therapeutically relevant phenotypic variants and/or disease predisposition. There are a number of well known polymorphisms that cause clinically important phenotypes; for example, the apoE2/3/4 variants are associated with different relative risk of Alzheimer's and other diseases (see Corder et al., Science 261(1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to oligonucleotide arrays has been shown to be an accurate and reliable method of simultaneously genotyping at least hundreds of SNPs; see Wang et al., Science, 280:1077 (1998); see also Schafer et al., Nature Biotechnology 16:33-39 (1998). The compositions of the present invention may easily be substituted for the arrays of the prior art.

15

20

25

There are a variety of particular techniques that are used to detect sequence, including mutations and SNPs. These include, but are not limited to, ligation based assays, cleavage based assays (mismatch and invasive cleavage such as Invader™), single base extension methods (see WO 92/15712, EP 0 371 437 B1, EP 0317 074 B1; Pastinen et al., Genome Res. 7:606-614 (1997); Syvänen, Clinica Chimica Acta 226:225-236 (1994); and WO 91/13075), and competitive probe analysis (e.g. competitive sequencing by hybridization; see below).

30

Oligonucleotide ligation amplification ("OLA", which is referred as the ligation chain reaction (LCR) when two-stranded reactions) involves the ligation of two smaller probes into a single long probe, using the target sequence as the template. See generally U.S. Patent Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; WO 97/31256 and WO 89/09835, all of which are incorporated by reference.

35

Invasive cleavage technology is based on structure-specific nucleases that cleave nucleic acids in a site-specific manner. Two probes are used: an "invader" probe and a "signalling" probe, that adjacently hybridize to a target sequence with a non-complementary overlap. The enzyme cleaves at the overlap due to its recognition of the "tail", and releases the "tail" with a label. This can then be detected. The Invader™ technology is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporated by reference.

An additional technique utilizes sequencing by hybridization. For example, sequencing by hybridization has been described (Drmanac et al., Genomics 4:114 (1989); Koster et al., Nature Biotechnology 14:1123 (1996); U.S. Patent Nos. 5,525,464; 5,202,231 and 5,695,940, among others, all of which are hereby expressly incorporated by reference in their entirety).

PCTs US98/21193, PCT US99/14387 and PCT US98/05025; WO98/50782; and U.S.S.N.s 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; all of which are expressly incorporated by reference, describe novel compositions utilizing substrates with microsphere arrays, which allow for novel detection methods of nucleic acid hybridization.

Accordingly, it is an object of the present invention to provide methods for determining the sequence of nucleic acids utilizing microsphere arrays.

SUMMARY OF THE INVENTION

In accordance with the above objects, the present invention provides methods of determining the identification of a nucleotide at a detection position in a target sequence. The methods comprise providing a hybridization complex comprising the target sequence and a capture probe covalently attached to a microsphere on a surface of a substrate. The methods comprise determining the nucleotide at the detection position. The hybridization complex can comprise the capture probe, a capture extender probe, and the target sequence. In addition, the target sequence may comprise an exogenous adapter sequences.

In an additional aspect, the method comprises contacting the microspheres with a plurality of detection probes each comprising a unique nucleotide at the readout position and a unique detectable label. The signal from at least one of the detectable labels is detected to identify the nucleotide at the detection position.

In a further aspect, the invention provides methods the target sequence comprises a first target domain directly 5' adjacent to the detection position. The hybridization complex comprises the target

sequence, a capture probe and an extension primer hybridized to the first target domain of the target sequence. The determination step comprises contacting the microspheres with a polymerase enzyme, and a plurality of NTPs each comprising a covalently attached detectable label, under conditions whereby if one of the NTPs basepairs with the base at the detection position, the extension primer is extended by the enzyme to incorporate the label. The base at the detection position is then identified.

In an additional aspect, the invention provides methods wherein the target sequence comprises a first target domain directly 5' adjacent to the detection position, wherein the capture probe serves an extension primer and is hybridized to the first target domain of the target sequence. The determination step comprises contacting the microspheres with a polymerase enzyme, and a plurality of NTPs each comprising a covalently attached detectable label, under conditions whereby if one of the NTPs basepairs with the base at the detection position, the extension primer is extended by the enzyme to incorporate the label. The base at the detection position is thus identified.

In a further aspect, the invention provides methods wherein the target sequence comprises (5' to 3'), a first target domain comprising an overlap domain comprising at least a nucleotide in the detection position and a second target domain contiguous with the detection position. The hybridization complex comprises a first probe hybridized to the first target domain, and a second probe hybridized to the second target domain. The second probe comprises a detection sequence that does not hybridize with the target sequence, and a detectable label. If the second probe comprises a base that is perfectly complementary to the detection position a cleavage structure is formed. The method further comprises contacting the hybridization complex with a cleavage enzyme that will cleave the detection sequence from the signalling probe and then forming an assay complex with the detection sequence, a capture probe covalently attached to a microsphere on a surface of a substrate, and at least one label. The base at the detection position is thus identified.

In an additional aspect, the invention provides methods of determining the identification of a nucleotide at a detection position in a target sequence comprising a first target domain comprising the detection position and a second target domain adjacent to the detection position. The method comprises hybridizing a first ligation probe to the first target domain, and hybridizing a second ligation probe to the second target domain. If the second ligation probe comprises a base that is perfectly complementary to the detection position a ligation structure is formed. A ligation enzyme is provided that will ligate the first and the second ligation probes to form a ligated probe. An assay complex is formed with the ligated probe, a capture probe covalently attached to a microsphere on a surface of a substrate, and at least one label. The base at the detection position is thus identified.

DETAILED DESCRIPTION OF THE FIGURES

Figures 1A, 1B, 1C, 1D and 1E schematically depict the use of readout probes. Figure 1A shows a "sandwich" format. Substrate **5** has a discrete site with a microsphere **10** comprising a capture probe **20** attached via a linker **15**. The target sequence **25** has a first domain that hybridizes to the capture probe **20** and a second domain comprising a detection position **30** that hybridizes to an readout probe **40** with readout position **35**. As will be appreciated by those in the art, Figure 1A depicts a single detection position; however, depending on the system, a plurality of different probes can hybridize to different target domains; hence n is an integer of 1 or greater. Figure 1B depicts the use of a capture probe **20** that also serves as an readout probe. Figure 1C depicts the use of an adapter probe **100** that binds to both the capture probe **20** and the target sequence **25**. As will be appreciated by those in the art, the figure depicts that the capture probe **20** and target sequence **25** bind adjacently and as such may be ligated; however, as will be appreciated by those in the art, there may be a "gap" of one or more nucleotides. Figure 1D depicts a solution based assay. Two readout probes **40**, each with a different readout position (**35** and **36**) and different labels (**45** and **46**) are added to target sequence **25** with detection position **35**, to form a hybridization complex with the match probe. This is added to the array; Figure 1D depicts the use of a capture probe **20** that directly hybridizes to a first domain of the target sequence, although other attachments may be done. Figure 1E depicts the direct attachment of the target sequence to the array.

Figures 2A, 2B, 2C, 2D, 2E and 2F depict preferred embodiments for SBE. Figure 2A depicts a "sandwich" assay, in which substrate **5** has a discrete site with a microsphere **10** comprising a capture probe **20** attached via a linker **15**. The target sequence **25** has a first domain that hybridizes to the capture probe **20** and a second domain comprising a detection position **30** that hybridizes to an extension primer **50**. As will be appreciated by those in the art, Figure 2A depicts a single detection position; however, depending on the system, a plurality of different primers can hybridize to different target domains; hence n is an integer of 1 or greater. In addition, the first domain of the target sequence may be an adapter sequence. Figure 2B depicts the use of a capture probe **20** that also serves as an extension primer. Figure 2C depicts the solution reaction. Figure 2D depicts the use of a capture extender probe **100**, that has a first domain that will hybridize to the capture probe **20** and a second domain that will hybridize to a first domain of the target sequence **25**.

Figures 3A, 3B, 3C, 3D and 3E depict some of the OLA embodiments of the reaction. Figure 3A depicts the solution reaction, wherein the target sequence **25** with a detection position **30** hybridizes to the first ligation probe **75** with readout position **35** and second probe **76** with a detectable label **45**. As will be appreciated by those in the art, the second ligation probe could also contain the readout position. The addition of a ligase forms a ligated probe **80**, that can then be added to the array with a

capture probe **20**. Figure 3B depicts an "on bead" assay, wherein the capture probe **20** serves as the first ligation probe. Figure 3C depicts a sandwich assay, using a capture probe **20** that hybridizes to a first portion of the target sequence **25** (which may be an endogenous sequence or an exogenous adapter sequence) and ligation probes **75** and **76** that hybridize to a second portion of the target sequence comprising the detection position **30**. Figure 3D depicts the use of a capture extender probe **100**. Figure 3E depicts a solution based assay with the use of an adapter sequence **110**.

Figures 4A, 4B and 4C depict the SPOLA reaction. In Figure 4A, two ligation probes are hybridized to a target sequence. As will be appreciated by those in the art, this system requires that the two ligation probes be attached at different ends, i.e. one at the 5' end and one at the 3' end. One of the ligation probes is attached via a cleavable linker. Upon formation of the assay complex and the addition of a ligase, the two probes will efficiently covalently couple the two ligation probes if perfect complementarity at the junction exists. In figure 4B, the resulting ligation difference between correctly matched probes and imperfect probes is shown. Figure 4C shows that subsequent cleavage of the cleavable linker produces a reactive group, in this case an amine, that may be subsequently labeled as outlined herein.

Figures 5A and 5B depict two cleavage reactions. Figure 5A depicts a loss of signal assay, wherein a label **45** is cleaved off due to the discrimination of the cleavage enzyme such as a restriction endonuclease. Figure 5B depicts the use of a quencher **46**.

Figure 6A, 6B, 6C, 6D, 6E and 6F depict the use of invasive cleavage to determine the identity of the nucleotide at the detection position. Figures 6A and 6B depict a loss of signal assay. Figure 6A depicts the invader probe **55** with readout position **35** hybridized to the target sequence **25** which is attached via a capture probe **20** to the surface. The signal probe **60** with readout position **35**, detectable label **45** and detection sequence **65** also binds to the target sequence **25**; the two probes form a cleavage structure. If the two readout positions **35** are capable of basepairing to the detection position **30** the addition of a structure-specific cleavage enzyme releases the detection sequence **65** and consequently the label **45**, leading to a loss of signal. Figure 6B is the same, except that the capture probe **20** also serves as the invader probe. Figure 6C depicts a solution reaction, wherein the signalling probe can comprise a capture tag **70** to facilitate the removal of uncleaved signal probes. The addition of the cleaved signal probe (e.g. the detection sequence **65**) with its associated label **45** results in detection. Figure 6D depicts a solution based assay using a label probe **120**. Figure 6E depicts a preferred embodiment of an invasive cleavage reaction that utilizes a fluorophore-quencher reaction. Figure 6E has the 3' end of the signal probe **60** is attached to the bead **10** and comprises a label **45** and a quencher **46**. Upon formation of the assay complex and subsequent cleavage, the quencher **46** is removed, leaving the fluorophore **45**.

Figures 7A, 7B, 7C and 7D depict assays based on the novel combination of competitive hybridization and extension. Figures 7A, 7B and 7C depict solution based assays. After hybridization of the extension probe **50** with a match base at the readout position **35**, an extension enzyme and dNTP is added, wherein the dNTP comprises a blocking moiety (to facilitate removal of unextended primers). Figure 7B depicts the same reaction with the use of an adapter sequence **90**; in this embodiment, the same adapter sequence **90** may be used for each readout probe for an allele. Figure 7C depicts the use of different adapter sequences **90** for each readout probe; in this embodiment, unreacted primers need not be removed, although they may be. Figure 7D depicts a solid phase reaction, wherein the dNTP added in the position adjacent to the readout position **35** is labeled.

Figures 8A and 8B depict assays based on the novel combination of invasive cleavage and ligation reactions. Figure 8A is a solution reaction, with the signalling probe **60** comprising a detection sequence **65** with a detectable label **45**. After hybridization with the target sequence **25** and cleavage, the free detection sequence can bind to an array (depicted herein as a bead array, although any nucleic acid array can be used), using a capture probe **20** and a template target sequence **26** for the ligation reaction. In the absence of ligation, the signalling probe is washed away. Figure 8B depicts a solid phase assay. In this embodiment, the 5' end of the signalling probe is attached to the array (again, depicted herein as a bead array, although any nucleic acid array can be used), and a blocking moiety is used at the 3' end. After cleavage, a free 3' end is generated, that can then be used for ligation using a template target **26**. As will be appreciated by those in the art, the orientation of this may be switched, such that the 3' end of the signalling probe **60** is attached, and a free 5' end is generated for the ligation reaction.

Figures 9A and 9B depict assays based on the novel combination of invasive cleavage and extension reactions. Figure 9A depicts an initial solution based assay, using a signalling probe with a blocked 3' end. After cleavage, the detection sequence can be added to an array and a template target added, followed by extension to add a detectable label. Alternatively, the extension can also happen in solution, using a template target **26**, followed by addition of the extended probe to the array. Figure 9B depicts the solid phase reaction; as above, either the 3' or the 5' end can be attached. By using a blocking moiety **47**, only the newly cleaved ends may be extended.

Figures 10A, 10B and 10C depict three configurations of the combination of ligation and extension ("Genetic Bit" analysis). Figure 10A depicts a reaction wherein the capture probe **20** and the extension probe serve as two ligation probes, and hybridize adjacently to the target sequence, such that an additional ligation step may be done. A labeled nucleotide is added at the readout position. Figure 10B depicts a preferred embodiment, wherein the ligation probes (one of which is the capture probe **20**) are separated by the detection position **30**. The addition of a labeled dNTP, extension

enzyme and ligase thus serve to detect the readout position. Figure 10C depicts the solution phase assay. As will be appreciated by those in the art, an extra level of specificity is added if the capture probe **20** spans the ligated probe **80**.

5

DETAILED DESCRIPTION OF THE INVENTION

10

This invention is directed to the detection (and optionally quantification) of differences or variations of sequences (e.g. SNPs) using bead arrays for detection of the differences. That is, the bead array serves as a platform on which a variety of techniques may be used to elucidate the nucleotide at the position of interest ("the detection position"). In general, the methods described herein relate to the detection of nucleotide substitutions, although as will be appreciated by those in the art, deletions, insertions, inversions, etc. may also be detected.

15

These techniques fall into five general categories: (1) techniques that rely on traditional hybridization methods that utilize the variation of stringency conditions (temperature, buffer conditions, etc.) to distinguish nucleotides at the detection position; (2) extension techniques that add a base ("the base") to basepair with the nucleotide at the detection position; (3) ligation techniques, that rely on the specificity of ligase enzymes (or, in some cases, on the specificity of chemical techniques), such that ligation reactions occur preferentially if perfect complementarity exists at the detection position; (4) cleavage techniques, that also rely on enzymatic or chemical specificity such that cleavage occurs preferentially if perfect complementarity exists; and (5) techniques that combine these methods.

20

25

Accordingly, the present invention provides compositions and methods for detecting the presence or absence of target nucleic acid sequences in a sample. As will be appreciated by those in the art, the sample solution may comprise any number of things, including, but not limited to, bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, anal and vaginal secretions, perspiration and semen, of virtually any organism, with mammalian samples being preferred and human samples being particularly preferred); environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples (i.e. in the case of nucleic acids, the sample may be the products of an amplification reaction, including both target and signal amplification as is generally described in "Detection of Nucleic Acid Amplification Reactions Using Bead Arrays", filed October 22, 1999 (no U.S.S.N. received yet; hereby incorporated by reference) such as PCR amplification reaction); purified samples, such as purified genomic DNA, RNA, proteins, etc.; raw samples (bacteria, virus, genomic DNA, etc.; As will be appreciated by those in the art, virtually any experimental manipulation may have been done on the sample.

30

35

The present invention provides compositions and methods for detecting the presence or absence of

target nucleic acid sequences in a sample. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for

5 example, phosphoramidate (Beaucage et al., *Tetrahedron* 49(10):1925 (1993) and references therein; Letsinger, J. *Org. Chem.* 35:3800 (1970); Sprinzl et al., *Eur. J. Biochem.* 81:579 (1977); Letsinger et al., *Nucl. Acids Res.* 14:3487 (1986); Sawai et al, *Chem. Lett.* 805 (1984), Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988); and Pauwels et al., *Chemica Scripta* 26:141 (1986)), phosphorothioate (Mag et al., *Nucleic Acids Res.* 19:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate

10 (Briu et al., *J. Am. Chem. Soc.* 111:2321 (1989), O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, *J. Am. Chem. Soc.* 114:1895 (1992); Meier et al., *Chem. Int. Ed. Engl.* 31:1008 (1992); Nielsen, *Nature*, 365:566 (1993); Carlsson et al., *Nature* 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with positive

15 backbones (Denpcy et al., *Proc. Natl. Acad. Sci. USA* 92:6097 (1995); non-ionic backbones (U.S. Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., *Angew. Chem. Intl. Ed. English* 30:423 (1991); Letsinger et al., *J. Am. Chem. Soc.* 110:4470 (1988); Letsinger et al., *Nucleoside & Nucleotide* 13:1597 (1994); Chapters 2 and 3, *ASC Symposium Series 580*, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker

20 et al., *Bioorganic & Medicinal Chem. Lett.* 4:395 (1994); Jeffs et al., *J. Biomolecular NMR* 34:17 (1994); *Tetrahedron Lett.* 37:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, *ASC Symposium Series 580*, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids

25 (see Jenkins et al., *Chem. Soc. Rev.* (1995) pp169-176). Several nucleic acid analogs are described in Rawls, *C & E News* June 2, 1997 page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of labels, alter the hybridization properties of the nucleic acids, or to increase the stability and half-life of such molecules in physiological environments.

30 As will be appreciated by those in the art, all of these nucleic acid analogs may find use in the present invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

35 Particularly preferred are peptide nucleic acids (PNA) which includes peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly

charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (T_m) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4°C drop in T_m for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9°C. This allows for better detection of mismatches. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribonucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc. A preferred embodiment utilizes isocytosine and isoguanine in nucleic acids designed to be complementary to other probes, rather than target sequences, as this reduces non-specific hybridization, as is generally described in U.S. Patent No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides as well as nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occurring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

The compositions and methods of the invention are directed to the detection of target sequences. The term "target sequence" or "target nucleic acid" or grammatical equivalents herein means a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. As is outlined herein, the target sequence may be a target sequence from a sample, or a derivative target such as a product of a reaction such as a detection sequence from an Invader™ reaction, a ligated probe from an OLA reaction, an extended probe from an SBE reaction, etc. It may be any length, with the understanding that longer sequences are more specific. As will be appreciated by those in the art, the complementary target sequence may take many forms. For example, it may be contained within a larger nucleic acid sequence, i.e. all or part of a gene or mRNA, a restriction fragment of a plasmid or genomic DNA, among others. As is outlined more fully below, probes are made to hybridize to target sequences to determine the presence or absence of the target sequence in a sample. Generally speaking, this term will be understood by those skilled in the art. The target sequence may also be comprised of different target domains; for example, a first target domain of the sample target sequence may hybridize to a capture probe, a second target domain may hybridize to a portion of a label probe, etc. The target domains may be adjacent or separated as indicated. Unless specified,

the terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target sequence. For example, assuming a 5'-3' orientation of the complementary target sequence, the first target domain may be located either 5' to the second domain, or 3' to the second domain. In addition, as will be appreciated by those in the art, the probes on the surface of the array (e.g. attached to the microspheres) may be attached in either orientation, either such that they have a free 3' end or a free 5' end; in some embodiments, the probes can be attached at one or more internal positions, or at both ends.

As is more fully outlined below, the target sequence comprises a position for which sequence information is desired, generally referred to herein as the "detection position" or "detection locus". In a preferred embodiment, the detection position is a single nucleotide, although in some embodiments, it may comprise a plurality of nucleotides, either contiguous with each other or separated by one or more nucleotides. By "plurality" as used herein is meant at least two. As used herein, the base which basepairs with a detection position base in a hybrid is termed a "readout position" or an "interrogation position".

In some embodiments, as is outlined herein, the target sequence may not be the sample target sequence but instead is a product of a reaction herein, sometimes referred to herein as a "secondary" or "derivative" target sequence. Thus, for example, in SBE, the extended primer may serve as the target sequence; similarly, in invasive cleavage variations, the cleaved detection sequence may serve as the target sequence.

If required, the target sequence is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, electroporation, etc., with purification and/or amplification as needed, as will be appreciated by those in the art. Suitable amplification techniques are outlined in "Detection of Nucleic Acid Amplification Reactions Using Bead Arrays", filed October 22, 1999 (no U.S.S.N. received yet) hereby expressly incorporated by reference.

Once prepared, the target sequence can be used in a variety of reactions for a variety of reasons. For example, in a preferred embodiment, genotyping reactions are done. Similarly, these reactions can also be used to detect the presence or absence of a target sequence. In addition, in any reaction, quantitation of the amount of a target sequence may be done. While the discussion below focuses on genotyping reactions, the discussion applies equally to detecting the presence of target sequences and/or their quantification.

Furthermore, as outlined below for each reaction, each of these techniques may be used in a solution based assay, wherein the reaction is done in solution and a reaction product is bound to the array for

subsequent detection, or in solid phase assays, where the reaction occurs on the surface and is detected.

These reactions are generally classified into 5 basic categories, as outlined below.

SIMPLE HYBRIDIZATION GENOTYPING

In a preferred embodiment, straight hybridization methods are used to elucidate the identity of the base at the detection position. Generally speaking, these techniques break down into two basic types of reactions: those that rely on competitive hybridization techniques, and those that discriminate using stringency parameters and combinations thereof.

Competitive hybridization

In a preferred embodiment, the use of competitive hybridization probes is done to elucidate either the identity of the nucleotide(s) at the detection position or the presence of a mismatch. For example, sequencing by hybridization has been described (Drmanac et al., Genomics 4:114 (1989); Koster et al., Nature Biotechnology 14:1123 (1996); U.S. Patent Nos. 5,525,464; 5,202,231 and 5,695,940, among others, all of which are hereby expressly incorporated by reference in their entirety).

It should be noted in this context that "mismatch" is a relative term and meant to indicate a difference in the identity of a base at a particular position, termed the "detection position" herein, between two sequences. In general, sequences that differ from wild type sequences are referred to as mismatches. However, particularly in the case of SNPs, what constitutes "wild type" may be difficult to determine as multiple alleles can be relatively frequently observed in the population, and thus "mismatch" in this context requires the artificial adoption of one sequence as a standard. Thus, for the purposes of this invention, sequences are referred to herein as "match" and "mismatch". Thus, the present invention may be used to detect substitutions, insertions or deletions as compared to a wild-type sequence.

In a preferred embodiment, a plurality of probes (sometimes referred to herein as "readout probes") are used to identify the base at the detection position. In this embodiment, each different readout probe comprises a different detection label (which, as outlined below, can be either a primary label or a secondary label) and a different base at the position that will hybridize to the detection position of the target sequence (herein referred to as the readout position) such that differential hybridization will occur. That is, all other parameters being equal, a perfectly complementary readout probe (a "match probe") will in general be more stable and have a slower off rate than a probe comprising a mismatch (a "mismatch probe") at any particular temperature. Accordingly, by using different readout probes, each with a different base at the readout position and each with a different label, the identification of

the base at the detection position is elucidated.

The readout probes comprise a detection label. By "detection label" or "detectable label" herein is meant a moiety that allows detection. This may be a primary label (which can be directly detected) or a secondary label (which is indirectly detected).

A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal labels; and c) colored or luminescent dyes. Preferred labels include chromophores or phosphors but are preferably fluorescent dyes. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, quantum dots (also referred to as "nanocrystals"), pyrene, Malachite green, stilbene, Lucifer Yellow, Cascade Blue™, Cy dyes (Cy3, Cy5, etc.), Texas Red, phycoerythrin, Bodipy, Alexa dyes and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference. In a preferred embodiment, the detection label used for competitive hybridization is a primary label.

In a preferred embodiment, the detectable label is a secondary label. A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a primary label for detection, can act on an additional product to generate a primary label (e.g. enzymes), or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE, OLA, invasive cleavage, etc. reactions; in addition, these techniques may be used with many of the other techniques described herein. Secondary labels include, but are not limited to, one of a binding partner pair; chemically modifiable moieties; nuclease inhibitors, enzymes such as horseradish peroxidase, alkaline phosphatases, luciferases, etc.

In a preferred embodiment, the secondary label is a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FABs, etc.)); proteins and small molecules, including biotin/streptavidin and digoxigenin and antibodies; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners, are also suitable binding pairs. Nucleic acid - nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP (or the probe) for incorporation into the extension primer. Preferred binding partner pairs include, but are not limited to, biotin (or imino-biotin) and streptavidin, digoxigenin and

Abs, and Prolinx™ reagents (see www.prolinxinc.com/ie4/home.html).

In a preferred embodiment, the binding partner pair comprises a primary detection label (attached to the NTP and therefore to the extended primer) and an antibody that will specifically bind to the primary detection label. By "specifically bind" herein is meant that the partners bind with specificity sufficient to differentiate between the pair and other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding. In some embodiments, the dissociation constants of the pair will be less than about 10^{-4} - 10^{-6} M⁻¹, with less than about 10^{-5} to 10^{-9} M⁻¹ being preferred and less than about 10^{-7} - 10^{-9} M⁻¹ being particularly preferred.

In addition, the secondary label can be a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid. Subsequently, primary labels, also comprising functional groups, may be added to these reactive groups. As is known in the art, this may be accomplished in a variety of ways. Preferred functional groups for attachment are amino groups, carboxy groups, oxo groups and thiol groups, with amino groups being particularly preferred. Using these functional groups, the primary labels can be attached using functional groups on the enzymes. For example, primary labels containing amino groups can be attached to secondary labels comprising amino groups, for example using linkers as are known in the art; for example, homo- or hetero-bifunctional linkers as are well known (see 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference).

Accordingly, a detectable label is incorporated into the readout probe. In a preferred embodiment, a set of readout probes are used, each comprising a different base at the readout position. In some embodiments, each readout probe comprises a different label, that is distinguishable from the others. For example, a first label may be used for probes comprising adenosine at the readout position, a second label may be used for probes comprising guanine at the readout position, etc. In a preferred embodiment, the length and sequence of each readout probe is identical except for the readout position, although this need not be true in all embodiments.

The number of readout probes used will vary depending on the end use of the assay. For example, many SNPs are biallelic, and thus two readout probes, each comprising an interrogation base that will basepair with one of the detection position bases. For sequencing, for example, for the discovery of SNPs, a set of four readout probes are used.

As will be appreciated by those in the art and additionally outlined below, this system can take on a number of different configurations, including a solution phase assay and a solid phase assay.

Solution phase assay

A solution phase assay that is followed by attaching the target sequence to an array is depicted in Figure 1D. In Figure 1D, a reaction with two different readout probes is shown. After the competitive hybridization has occurred, the target sequence is added to the array, which may take on several configurations, outlined below.

Solid phase assay

In a preferred embodiment, the competition reaction is done on the array. This system may take on several configurations.

In a preferred embodiment, a sandwich assay of sorts is used. In this embodiment, the bead comprises a capture probe that will hybridize to a first target domain of a target sequence, and the readout probe will hybridize to a second target domain, as is generally depicted in Figure 1A. In this embodiment, the first target domain may be either unique to the target, or may be an exogenous adapter sequence added to the target sequence as outlined below, for example through the use of PCR reactions. Similarly, a sandwich assay that utilizes a capture extender probe, as described below, to attach the target sequence to the array is depicted in Figure 1C.

Alternatively, the capture probe itself can be the readout probe as is shown in Figure 1B; that is, a plurality of microspheres are used, each comprising a capture probe that has a different base at the readout position. In general, the target sequence then hybridizes preferentially to the capture probe most closely matched. In this embodiment, either the target sequence itself is labeled (for example, it may be the product of an amplification reaction) or a label probe may bind to the target sequence at a domain remote from the detection position. In this embodiment, since it is the location on the array that serves to identify the base at the detection position, different labels are not required.

In a further embodiment, the target sequence itself is attached to the array, as generally depicted for bead arrays in Figure 1E and described below.

Stringency Variation

In a preferred embodiment, sensitivity to variations in stringency parameters are used to determine either the identity of the nucleotide(s) at the detection position or the presence of a mismatch. As a preliminary matter, the use of different stringency conditions such as variations in temperature and buffer composition to determine the presence or absence of mismatches in double stranded hybrids comprising a single stranded target sequence and a probe is well known.

With particular regard to temperature, as is known in the art, differences in the number of hydrogen

bonds as a function of basepairing between perfect matches and mismatches can be exploited as a result of their different T_m s (the temperature at which 50% of the hybrid is denatured). Accordingly, a hybrid comprising perfect complementarity will melt at a higher temperature than one comprising at least one mismatch, all other parameters being equal. (It should be noted that for the purposes of the discussion herein, all other parameters (i.e. length of the hybrid, nature of the backbone (i.e. naturally occurring or nucleic acid analog), the assay solution composition and the composition of the bases, including G-C content are kept constant). However, as will be appreciated by those in the art, these factors may be varied as well, and then taken into account.)

In general, as outlined herein, high stringency conditions are those that result in perfect matches remaining in hybridization complexes, while imperfect matches melt off. Similarly, low stringency conditions are those that allow the formation of hybridization complexes with both perfect and imperfect matches. High stringency conditions are known in the art; see for example Maniatis et al., *Molecular Cloning: A Laboratory Manual*, 2d Edition, 1989, and *Short Protocols in Molecular Biology*, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, *supra*, and Tijssen, *supra*.

As will be appreciated by those in the art, mismatch detection using temperature may proceed in a variety of ways, and is similar to the use of readout probes as outlined above. Again, as outlined above, a plurality of readout probes may be used in a sandwich format; in this embodiment, all the probes may bind at permissive, low temperatures (temperatures below the T_m of the mismatch); however, repeating the assay at a higher temperature (above the T_m of the mismatch) only the

perfectly matched probe may bind. Thus, this system may be run with readout probes with different detectable labels, as outlined above. Alternatively, a single probe may be used to query whether a particular base is present.

5 Alternatively, as described above, the capture probe may serve as the readout probe; in this embodiment, a single label may be used on the target; at temperatures above the T_m of the mismatch, only signals from perfect matches will be seen, as the mismatch target will melt off.

10 Similarly, variations in buffer composition may be used to elucidate the presence or absence of a mismatch at the detection position. Suitable conditions include, but are not limited to, formamide concentration. Thus, for example, "low" or "permissive" stringency conditions include formamide concentrations of 0 to 10%, while "high" or "stringent" conditions utilize formamide concentrations of $\geq 40\%$. Low stringency conditions include NaCl concentrations of ≥ 1 M, and high stringency conditions include concentrations of ≤ 0.3 M. Furthermore, low stringency conditions include $MgCl_2$
15 concentrations of ≥ 10 mM, moderate stringency as 1-10 mM, and high stringency conditions include concentrations of ≤ 1 mM.

In this embodiment, as for temperature, a plurality of readout probes may be used, with different bases in the readout position (and optionally different labels). Running the assays under the permissive
20 conditions and repeating under stringent conditions will allow the elucidation of the base at the detection position.

In one embodiment, the probes used as readout probes are "Molecular Beacon" probes as are generally described in Whitcombe et al., Nature Biotechnology 17:804 (1999), hereby incorporated by
25 reference. As is known in the art, Molecular Beacon probes form "hairpin" type structures, with a fluorescent label on one end and a quencher on the other. In the absence of the target sequence, the ends of the hairpin hybridize, causing quenching of the label. In the presence of a target sequence, the hairpin structure is lost in favor of target sequence binding, resulting in a loss of quenching and thus an increase in signal.

30 In one embodiment, the Molecular Beacon probes can be the capture probes as outlined herein for readout probes. For example, different beads comprising labeled Molecular Beacon probes (and different bases at the readout position) are made optionally they comprise different labels.

35 Alternatively, since Molecular Beacon probes do not have an appreciable signal in the absence of their target sequence, all four probes (if a set of four different bases with is used) differently labelled are attached to a single bead.

EXTENSION GENOTYPING

In this embodiment, any number of techniques are used to add a nucleotide to the readout position of a probe hybridized to the target sequence adjacent to the detection position. By relying on enzymatic specificity, preferentially a perfectly complementary base is added. All of these methods rely on the enzymatic incorporation of nucleotides at the detection position. This may be done using chain terminating dNTPs, such that only a single base is incorporated (e.g. single base extension methods), or under conditions that only a single type of nucleotide is added followed by identification of the added nucleotide (extension and pyrosequencing techniques).

Single Base Extension

In a preferred embodiment, single base extension (SBE; sometimes referred to as "minisequencing") is used to determine the identity of the base at the detection position. Briefly, SBE is a technique that utilizes an extension primer that hybridizes to the target nucleic acid immediately adjacent to the detection position. A polymerase (generally a DNA polymerase) is used to extend the 3' end of the primer with a nucleotide analog labeled a detection label as described herein. Based on the fidelity of the enzyme, a nucleotide is only incorporated into the readout position of the growing nucleic acid strand if it is perfectly complementary to the base in the target strand at the detection position. The nucleotide may be derivatized such that no further extensions can occur, so only a single nucleotide is added. Once the labeled nucleotide is added, detection of the label proceeds as outlined herein. See generally Sylvanen et al., Genomics 8:684-692 (1990); U.S. Patent Nos. 5,846,710 and 5,888,819; Pastinen et al., Genomics Res. 7(6):606-614 (1997); all of which are expressly incorporated herein by reference.

The reaction is initiated by introducing the assay complex comprising the target sequence (i.e. the array) to a solution comprising a first nucleotide. By "nucleotide" in this context herein is meant a deoxynucleoside-triphosphate (also called deoxynucleotides or dNTPs, e.g. dATP, dTTP, dCTP and dGTP). In general, the nucleotides comprise a detectable label, which may be either a primary or a secondary label. In addition, the nucleotides may be nucleotide analogs, depending on the configuration of the system. For example, if the dNTPs are added in sequential reactions, such that only a single type of dNTP can be added, the nucleotides need not be chain terminating. In addition, in this embodiment, the dNTPs may all comprise the same type of label.

Alternatively, if the reaction comprises more than one dNTP, the dNTPs should be chain terminating, that is, they have a blocking or protecting group at the 3' position such that no further dNTPs may be added by the enzyme. As will be appreciated by those in the art, any number of nucleotide analogs may be used, as long as a polymerase enzyme will still incorporate the nucleotide at the readout position. Preferred embodiments utilize dideoxy-triphosphate nucleotides (ddNTPs) and halogenated

dNTPs. Generally, a set of nucleotides comprising ddATP, ddCTP, ddGTP and ddTTP is used, each with a different detectable label, although as outlined herein, this may not be required.

In a preferred embodiment, the nucleotide analogs comprise a detectable label, which can be either a primary or secondary detectable label. Preferred primary labels are those outlined above for interrogation labels. However, the enzymatic incorporation of nucleotides comprising fluorophores is may be poor under many conditions; accordingly, a preferred embodiment utilizes secondary detectable labels. In addition, as outlined below, the use of secondary labels may also facilitate the removal of unextended probes.

In addition, as will be appreciated by those in the art, the single base extension reactions of the present invention allow the precise incorporation of modified bases into a growing nucleic acid strand. Thus, any number of modified nucleotides may be incorporated for any number of reasons, including probing structure-function relationships (e.g. DNA:DNA or DNA:protein interactions), cleaving the nucleic acid, crosslinking the nucleic acid, incorporate mismatches, etc.

In addition to a first nucleotide, the solution also comprises an extension enzyme, generally a DNA polymerase. Suitable DNA polymerases include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase. If the NTP is complementary to the base of the detection position of the target sequence, which is adjacent to the extension primer, the extension enzyme will add it to the extension primer at the readout position. Thus, the extension primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strand". If desired, the temperature of the reaction can be adjusted (or cycled) such that amplification occurs, generating a plurality of modified primers.

As will be appreciated by those in the art, the configuration of the SBE system can take on several forms.

Solution phase assay

As for the OLA reaction described below, the reaction may be done in solution, and then the newly synthesized strands, with the base-specific detectable labels, can be detected. For example, they can be directly hybridized to capture probes that are complementary to the extension primers, and the presence of the label is then detected. This is schematically depicted in Figure 2C. As will be appreciated by those in the art, a preferred embodiment utilizes four different detectable labels, i.e. one for each base, such that upon hybridization to the capture probe on the array, the identification of the base can be done isothermally. Thus, Figure 2C depicts the readout position 35 as not

neccessarily hybridizing to the capture probe.

In a preferred embodiment, adapter sequences can be used in a solution format. In this embodiment, a single label can be used with a set of four separate primer extension reactions. In this embodiment, the extension reaction is done in solution; each reaction comprises a different dNTP with the label. For each locus genotyped, a set of four different extension primers are used, each with a portion that will hybridize to the target sequence, a different readout base and each with a different adapter sequence of 15-40 bases, as is more fully outlined below. After the primer extension reaction is complete, the four separate reactions are pooled and hybridized to an array comprising complementary probes to the adapter sequences. A genotype is derived by comparing the probe intensities of the four different hybridized adapter sequences corresponding to a give locus.

In addition, since unextended primers do not comprise labels, the unextended primers need not be removed. However, they may be, if desired, as outlined below; for example, if a large excess of primers are used, there may not be sufficient signal from the extended primers competing for binding to the surface.

Alternatively, one of skill in the art could use a single label and temperature to determine the identity of the base; that is, the readout position of the extension primer hybridizes to a position on the capture probe. However, since the three mismatches will have lower Tms than the perfect match, the use of temperature could elucidate the identity of the detection position base.

Solid phase assay

Alternatively, the reaction may be done on a surface by capturing the target sequence and then running the SBE reaction, in a sandwich type format schematically depicted in Figure 2A. In this embodiment, the capture probe hybridizes to a first domain of the target sequence (which can be endogeneous or an exogeneous adapter sequence added during an amplification reaction), and the extension primer hybridizes to a second target domain immediately adjacent to the detection position. The addition of the enzyme and the required NTPs results in the addition of the interrogation base. In this embodiment, each NTP must have a unique label. Alternatively, each NTP reaction may be done sequentially on a different array.

Furthermore, as is more fully outlined below and depicted in Figure 2D, capture extender probes can be used to attach the target sequence to the bead. In this embodiment, the hybridization complex comprises the capture probe, the target sequence and the adapter sequence.

Similarly, the capture probe itself can be used as the extension probe, with its terminus being directly

adjacent to the detection position. This is schematically depicted in Figure 2B. Upon the addition of the target sequence and the SBE reagents, the modified primer is formed comprising a detectable label, and then detected. Again, as for the solution based reaction, each NTP must have a unique label, the reactions must proceed sequentially, or different arrays must be used.

In addition, as outlined herein, the target sequence may be directly attached to the array; the extension primer hybridizes to it and the reaction proceeds.

Variations on this is shown in Figures 2E and 2F, where the the capture probe and the extension probe adjacently hybridize to the target sequence. Either before or after extension of the extension probe, a ligation step may be used to attach the capture and extension probes together for stability. These are further described below as combination assays.

As will be appreciated by those in the art, the determination of the base at the detection position can proceed in several ways. In a preferred embodiment, the reaction is run with all four nucleotides (assuming all four nucleotides are required), each with a different label, as is generally outlined herein. Alternatively, a single label is used, by using four reactions: this may be done either by using a single substrate and sequential reactions, or by using four arrays. For example, dATP can be added to the assay complex, and the generation of a signal evaluated; the dATP can be removed and dTTP added, etc. Alternatively, four arrays can be used; the first is reacted with dATP, the second with dTTP, etc., and the presence or absence of a signal evaluated.

Alternatively, ratiometric analysis can be done; for example, two labels, "A" and "B", on two substrates (e.g. two arrays) can be done. In this embodiment, two sets of primer extension reactions are performed, each on two arrays, with each reaction containing a complete set of four chain terminating NTPs. The first reaction contains two "A" labeled nucleotides and two "B" labeled nucleotides (for example, A and C may be "A" labeled, and G and T may be "B" labeled). The second reaction also contains the two labels, but switched; for example, A and G are "A" labeled and T and C are "B" labeled. This reaction composition allows a biallelic marker to be ratiometrically scored; that is, the intensity of the two labels in two different "color" channels on a single substrate is compared, using data from a set of two hybridized arrays. For instance, if the marker is A/G, then the first reaction on the first array is used to calculate a ratiometric genotyping score; if the marker is A/C, then the second reaction on the second array is used for the calculation; if the marker is G/T, then the second array is used, etc. This concept can be applied to all possible biallelic marker combinations. "Scoring" a genotype using a single fiber ratiometric score allows a much more robust genotyping than scoring a genotype using a comparison of absolute or normalized intensities between two different arrays.

Removal of unextended primers

In a preferred embodiment, for both SBE as well as a number of other reactions outlined herein, it is desirable to remove the unextended or unreacted primers from the assay mixture, and particularly from the array, as unextended primers will compete with the extended (labeled) primers in binding to capture probes, thereby diminishing the signal. The concentration of the unextended primers relative to the extended primer may be relatively high, since a large excess of primer is usually required to generate efficient primer annealing. Accordingly, a number of different techniques may be used to facilitate the removal of unextended primers. These generally include methods based on removal of unreacted primers by binding to a solid support, protecting the reacted primers and degrading the unextended ones, and separating the unreacted and reacted primers. While the discussion below applies specifically to SBE, these techniques may be used in any of the methods described herein.

Solid phase removal

In a preferred embodiment, the NTPs (or, in the case of other methods, one or more of the probes) comprise a secondary detectable label that can be used to separate extended and non-extended primers. As outlined above, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable). Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE, OLA, invasive cleavage, etc. reactions; in addition, these techniques may be used with many of the other techniques described herein.

In a preferred embodiment, the secondary label is a one of a binding partner pair. For example, a preferred embodiment utilizes binding partner pairs comprising biotin or imino-biotin and streptavidin. Imino-biotin is particularly preferred when the methods require the later separation of the pair, as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95°C).

This may also be accomplished using chemically modifiable secondary labels. That is, in a preferred embodiment, the secondary label is a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid. These functional groups are then used to remove the reacted primers, for example by attaching the reacted primers to a solid support, as outlined below, followed by a cleavage reaction and addition to the array.

In this embodiment, it is preferred that the other half of the binding pair is attached to a solid support. In this embodiment, the solid support may be any as described herein for substrates and microspheres, and the form is preferably microspheres as well; for example, a preferred embodiment utilizes magnetic beads that can be easily introduced to the sample and easily removed, although any affinity chromatography formats may be used as well. Standard methods are used to attach the

binding partner to the solid support, and can include direct or indirect attachment methods. For example, biotin labeled antibodies to fluorophores can be attached to streptavidin coated magnetic beads.

Thus, in this embodiment, the extended primers comprise a binding member that is contacted with its binding partner under conditions wherein the extended primers are separated from the unextended primers. These extended primers can then be added to the array comprising capture probes as described herein.

Protection and degradation

In this embodiment, the dNTPs that are added during the reaction confer protection from degradation (whether chemical or enzymatic). Thus, after the assay, the degradation components are added, and unreacted primers are degraded, leaving only the reacted primers. Labeled protecting groups are particularly preferred; for example, 3'-substituted-2'-dNTPs can contain anthranilic derivatives that are fluorescent (with alkali or enzymatic treatment for removal of the protecting group).

In a preferred embodiment, the secondary label is a nuclease inhibitor, such as thiol NTPs. In this embodiment, the chain-terminating NTPs are chosen to render extended primers resistant to nucleases, such as 3'-exonucleases. Addition of an exonuclease will digest the non-extended primers leaving only the extended primers to bind to the capture probes on the array. This may also be done with OLA, wherein the ligated probe will be protected but the unprotected ligation probe will be digested.

In this embodiment, suitable 3'-exonucleases include, but are not limited to, exo I, exo III, exo VII, and 3'-5' exophosphodiesterases.

Alternatively, an 3' exonuclease may be added to a mixture of 3' labeled biotin/streptavidin; only the unreacted oligonucleotides will be degraded. Following exonuclease treatment, the exonuclease and the streptavidin can be degraded using a protease such as proteinase K. The surviving nucleic acids (i.e. those that were biotinylated) are then hybridized to the array.

Separation systems

The use of secondary label systems (and even some primary label systems) can be used to separate unreacted and reacted probes; for example, the addition of streptavidin to a nucleic acid greatly increases its size, as well as changes its physical properties, to allow more efficient separation techniques. For example, the mixtures can be size fractionated by exclusion chromatography, affinity chromatography, filtration or differential precipitation.

Non-terminated extension

In a preferred embodiment, methods of adding a single base are used that do not rely on chain termination. That is, similar to SBE, enzymatic reactions that utilize dNTPs and polymerases can be used; however, rather than use chain terminating dNTPs, regular dNTPs are used. This method relies on a time-resolved basis of detection; only one type of base is added during the reaction. Thus, for example, four different reactions each containing one of the dNTPs can be done; this is generally accomplished by using four different substrates, although as will be appreciated by those in the art, not all four reactions need occur to identify the nucleotide at a detection position. In this embodiment, the signals from single additions can be compared to those from multiple additions; that is, the addition of a single ATP can be distinguished on the basis of signal intensity from the addition of two or three ATPs. These reactions are accomplished as outlined above for SBE, using extension primers and polymerases; again, one label or four different labels can be used, although as outlined herein, the different NTPs must be added sequentially.

A preferred method of extension in this embodiment is pyrosequencing.

Pyrosequencing

Pyrosequencing is an extension method that can be used to add one or more nucleotides to the detection position(s); it is very similar to SBE except that chain terminating NTPs need not be used (although they may be). Pyrosequencing relies on the detection of a reaction product, PPi, produced during the addition of an NTP to a growing oligonucleotide chain, rather than on a label attached to the nucleotide. One molecule of PPi is produced per dNTP added to the extension primer. That is, by running sequential reactions with each of the nucleotides, and monitoring the reaction products, the identity of the added base is determined.

The release of pyrophosphate (PPi) during the DNA polymerase reaction can be quantitatively measured by many different methods and a number of enzymatic methods have been described; see Reeves et al., Anal. Biochem. 28:282 (1969); Guillory et al., Anal. Biochem. 39:170 (1971); Johnson et al., Anal. Biochem. 15:273 (1968); Cook et al., Anal. Biochem. 91:557 (1978); Drake et al., Anal. Biochem. 94:117 (1979); WO93/23564; WO 98/28440; WO98/13523; Nyren et al., Anal. Biochem. 151:504 (1985); all of which are incorporated by reference. The latter method allows continuous monitoring of PPi and has been termed ELIDA (Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay). A preferred embodiment utilizes any method which can result in the generation of an optical signal, with preferred embodiments utilizing the generation of a chemiluminescent or fluorescent signal.

A preferred method monitors the creation of PPi by the conversion of PPi to ATP by the enzyme

sulfurylase, and the subsequent production of visible light by firefly luciferase (see Ronaghi et al., Science 281:363 (1998), incorporated by reference). In this method, the four deoxynucleotides (dATP, dGTP, dCTP and dTTP; collectively dNTPs) are added stepwise to a partial duplex comprising a sequencing primer hybridized to a single stranded DNA template and incubated with DNA polymerase, ATP sulfurylase, luciferase, and optionally a nucleotide-degrading enzyme such as apyrase. A dNTP is only incorporated into the growing DNA strand if it is complementary to the base in the template strand. The synthesis of DNA is accompanied by the release of PPi equal in molarity to the incorporated dNTP. The PPi is converted to ATP and the light generated by the luciferase is directly proportional to the amount of ATP. In some cases the unincorporated dNTPs and the produced ATP are degraded between each cycle by the nucleotide degrading enzyme.

Accordingly, a preferred embodiment of the methods of the invention is as follows. A substrate comprising microspheres containing the target sequences and extension primers, forming hybridization complexes, is dipped or contacted with a reaction chamber or well comprising a single type of dNTP, an extension enzyme, and the reagents and enzymes necessary to detect PPi. If the dNTP is complementary to the base of the target portion of the target sequence adjacent to the extension primer, the dNTP is added, releasing PPi and generating detectable light, which is detected as generally described in U.S.S.N.s 09/151,877 and 09/189,543, and PCT US98/09163, all of which are hereby incorporated by reference. If the dNTP is not complementary, no detectable signal results. The substrate is then contacted with a second reaction chamber comprising a different dNTP and the additional components of the assay. This process is repeated if the identity of a base at a second detection position is desirable.

In a preferred embodiment, washing steps, i.e. the use of washing chambers, may be done in between the dNTP reaction chambers, as required. These washing chambers may optionally comprise a nucleotide-degrading enzyme, to remove any unreacted dNTP and decreasing the background signal, as is described in WO 98/28440, incorporated herein by reference.

As will be appreciated by those in the art, the system can be configured in a variety of ways, including both a linear progression or a circular one; for example, four arrays may be used that each can dip into one of four reaction chambers arrayed in a circular pattern. Each cycle of sequencing and reading is followed by a 90 degree rotation, so that each substrate then dips into the next reaction well.

In a preferred embodiment, one or more internal control sequences are used. That is, at least one microsphere in the array comprises a known sequence that can be used to verify that the reactions are proceeding correctly. In a preferred embodiment, at least four control sequences are used, each of which has a different nucleotide at each position: the first control sequence will have an adenosine

at position 1, the second will have a cytosine, the third a guanosine, and the fourth a thymidine, thus ensuring that at least one control sequence is "lighting up" at each step to serve as an internal control.

As for simple extension and SBE, the pyrosequencing systems may be configured in a variety of ways; for example, the target sequence may be attached to the bead in a variety of ways, including direct attachment of the target sequence; the use of a capture probe with a separate extension probe; the use of a capture extender probe, a capture probe and a separate extension probe; the use of adapter sequences in the target sequence with capture and extension probes; and the use of a capture probe that also serves as the extension probe.

One additional benefit of pyrosequencing for genotyping purposes is that since the reaction does not rely on the incorporation of labels into a growing chain, the unreacted extension primers need not be removed.

Allelic PCR

In a preferred embodiment, the method used to detect the base at the detection position is allelic PCR, referred to herein as "aPCR". As described in Newton et al., Nucl. Acid Res. 17:2503 (1989), hereby expressly incorporated by reference, allelic PCR allows single base discrimination based on the fact that the PCR reaction does not proceed well if the terminal 3'-nucleotide is mismatched, assuming the DNA polymerase being used lacks a 3'-exonuclease proofreading activity. Accordingly, the identification of the base proceeds by using allelic PCR primers (sometimes referred to herein as aPCR primers) that have readout positions at their 3' ends. Thus the target sequence comprises a first domain comprising at its 5' end a detection position.

In general, aPCR may be briefly described as follows. A double stranded target nucleic acid is denatured, generally by raising the temperature, and then cooled in the presence of an excess of a aPCR primer, which then hybridizes to the first target strand. If the readout position of the aPCR primer basepairs correctly with the detection position of the target sequence, a DNA polymerase (again, that lacks 3'-exonuclease activity) then acts to extend the primer with dNTPs, resulting in the synthesis of a new strand forming a hybridization complex. The sample is then heated again, to disassociate the hybridization complex, and the process is repeated. By using a second PCR primer for the complementary target strand, rapid and exponential amplification occurs. Thus aPCR steps are denaturation, annealing and extension. The particulars of aPCR are well known, and include the use of a thermostable polymerase such as Taq I polymerase and thermal cycling.

Accordingly, the aPCR reaction requires at least one aPCR primer, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

Furthermore, the aPCR reaction may be run as a competition assay of sorts. For example, for biallelic SNPs, a first aPCR primer comprising a first base at the readout position and a first label, and a second aPCR primer comprising a different base at the readout position and a second label, may be used. The PCR primer for the other strand is the same. The examination of the ratio of the two colors can serve to identify the base at the detection position.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially complementary to the extended part of the primer; that is, unextended primers will not bind to the capture probes.

LIGATION TECHNIQUES FOR GENOTYPING

In this embodiment, the readout of the base at the detection position proceeds using a ligase. In this embodiment, it is the specificity of the ligase which is the basis of the genotyping; that is, ligases generally require that the 5' and 3' ends of the ligation probes have perfect complementarity to the target for ligation to occur.

In a preferred embodiment, the identity of the base at the detection position proceeds utilizing the OLA, as is generally depicted in Figure 3. The method can be run at least two different ways; in a first embodiment, only one strand of a target sequence is used as a template for ligation; alternatively, both strands may be used; the latter is generally referred to as Ligation Chain Reaction or LCR. See generally U.S. Patent Nos. 5,185,243 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, and U.S.S.N.s 60/078,102 and 60/073,011, all of which are incorporated by reference.

This method is based on the fact that two probes can be preferentially ligated together, if they are hybridized to a target strand and if perfect complementarity exists at the two bases being ligated together. Thus, in this embodiment, the target sequence comprises a contiguous first target domain comprising the detection position and a second target domain adjacent to the detection position. That is, the detection position is "between" the rest of the first target domain and the second target domain. A first ligation probe is hybridized to the first target domain and a second ligation probe is hybridized to the second target domain. If the first ligation probe has a base perfectly complementary to the detection position base, and the adjacent base on the second probe has perfect complementarity to its position, a ligation structure is formed such that the two probes can be ligated together to form a

ligated probe. If this complementarity does not exist, no ligation structure is formed and the probes are not ligated together to an appreciable degree. This may be done using heat cycling, to allow the ligated probe to be denatured off the target sequence such that it may serve as a template for further reactions. In addition, as is more fully outlined below, this method may also be done using ligation probes that are separated by one or more nucleotides, if dNTPs and a polymerase are added (this is sometimes referred to as "Genetic Bit" analysis).

In a preferred embodiment, LCR is done for two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set (i.e. third and fourth primer probe nucleic acids) for the other strand of the target. In a preferred embodiment, the first and third probes will hybridize, and the second and fourth probes will hybridize, such that amplification can occur. That is, when the first and second probes have been attached, the ligated probe can now be used as a template, in addition to the second target sequence, for the attachment of the third and fourth probes. Similarly, the ligated third and fourth probes will serve as a template for the attachment of the first and second probes, in addition to the first target strand. In this way, an exponential, rather than just a linear, amplification can occur.

As will be appreciated by those in the art, the ligation product can be detected in a variety of ways. Preferably, detection is accomplished by removing the unligated labeled probe from the reaction before application to a capture probe. In one embodiment, the unligated probes are removed by digesting 3' non-protected oligonucleotides with a 3' exonuclease, such as, exonuclease I. The ligation products are protected from exo I digestion by including, for example, the use of a number of sequential phosphorothioate residues at their 3' terminus (for example at least four), thereby, rendering them resistant to exonuclease digestion. The unligated detection oligonucleotides are not protected and are digested.

As for most or all of the methods described herein, the assay can take on a solution-based form or a solid-phase form.

Solution based OLA

In a preferred embodiment, as shown in Figure 3A, the ligation reaction is run in solution. In this embodiment, only one of the primers carries a detectable label, e.g. the first ligation probe, and the capture probe on the bead is substantially complementary to the other probe, e.g. the second ligation probe. In this way, unextended labeled ligation primers will not interfere with the assay. This substantially reduces or eliminates false signal generated by the optically-labeled 3' primers.

In addition, a solution-based OLA assay that utilizes adapter sequences may be done. In this embodiment, rather than have the target sequence comprise the adapter sequences, one of the ligation probes comprises the adapter sequence. This facilitates the creation of "universal arrays". For example, as depicted in Figure 3E, the first ligation probe has an adapter sequence that is used to attach the ligated probe to the array.

Solid phase based OLA

Alternatively, the target nucleic acid is immobilized on a solid-phase surface. The OLA assay is performed and unligated oligonucleotides are removed by washing under appropriate stringency to remove unligated oligonucleotides and thus the label. For example, as depicted in Figure 3B, the capture probe can comprise one of the ligation probes. Similarly, Figures 3C and 3D depict alternative attachments.

Again, as outlined above, the detection of the OLA reaction can also occur directly, in the case where one or both of the primers comprises at least one detectable label, or indirectly, using sandwich assays, through the use of additional probes; that is, the ligated probes can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc.

Again, as outlined above for SBE, unreacted ligation primers may be removed from the mixture as needed. For example, the first ligation probe may comprise the label (either a primary or secondary label) and the second may be blocked at its 3' end with an exonuclease blocking moiety; after ligation and the introduction of the nuclease, the labeled ligation probe will be digested, leaving the ligation product and the second probe; however, since the second probe is unlabeled, it is effectively silent in the assay. Similarly, the second probe may comprise a binding partner used to pull out the ligated probes, leaving unligated labeled ligation probes behind. The binding pair is then disassociated and added to the array.

Solid Phase Oligonucleotide Ligation Assay (SPOLA)

In a preferred embodiment, a novel method of OLA is used, termed herein "solid phase oligonucleotide assay", or "SPOLA". In this embodiment, the ligation probes are both attached to the same site on the surface of the array (e.g. when microsphere arrays are used, to the same bead), one at its 5' end (the "upstream probe") and one at its 3' end (the "downstream probe"), as is generally depicted in Figure 4. This may be done as is will be appreciated by those in the art. At least one of the probes is attached via a cleavable linker, that upon cleavage, forms a reactive moiety. If ligation occurs, the reactive moiety remains associated with the surface; but if no ligation occurs, due to a mismatch, the reactive moiety is free in solution to diffuse away from the surface of the array. The reactive moiety is then

used to add a detectable label.

Generally, as will be appreciated by those in the art, cleavage of the cleavable linker should result in asymmetrical products; i.e. one of the "ends" should be reactive, and the other should not, with the configuration of the system such that the reactive moiety remains associated with the surface if ligation occurred. Thus, for example, amino acids or succinate esters can be cleaved either enzymatically (via peptidases (aminopeptidase and carboxypeptidase) or proteases) or chemically (acid/base hydrolysis) to produce an amine and a carboxyl group. One of these groups can then be used to add a detectable label, as will be appreciated by those in the art and discussed herein.

Padlock probe ligation

In a preferred embodiment, the ligation probes are specialized probes called "padlock probes". Nilsson et al, 1994, Science 265:2085. These probes have a first ligation domain that is identical to a first ligation probe, in that it hybridizes to a first target sequence domain, and a second ligation domain, identical to the second ligation probe, that hybridizes to an adjacent target sequence domain. Again, as for OLA, the detection position can be either at the 3' end of the first ligation domain or at the 5' end of the second ligation domain. However, the two ligation domains are connected by a linker, frequently nucleic acid. The configuration of the system is such that upon ligation of the first and second ligation domains of the padlock probe, the probe forms a circular probe, and forms a complex with the target sequence wherein the target sequence is "inserted" into the loop of the circle.

In this embodiment, the unligated probes may be removed through degradation (for example, through a nuclease), as there are no "free ends" in the ligated probe.

CLEAVAGE TECHNIQUES FOR GENOTYPING

In a preferred embodiment, the specificity for genotyping is provided by a cleavage enzyme. There are a variety of enzymes known to cleave at specific sites, either based on sequence specificity, such as restriction endonucleases, or using structural specificity, such as is done through the use of invasive cleavage technology.

ENDONUCLEASE TECHNIQUES

In a preferred embodiment, enzymes that rely on sequence specificity are used. In general, these systems rely on the cleavage of double stranded sequence containing a specific sequence recognized by a nuclease, preferably an endonuclease.

These systems may work in a variety of ways, as is generally depicted in Figure 6. In one

embodiment (Figure 6A), a labeled readout probe (generally attached to a bead of the array) is used; the binding of the target sequence forms a double stranded sequence that a restriction endonuclease can then recognize and cleave, if the correct sequence is present. An enzyme resulting in "sticky ends" is shown in Figure 6A. The cleavage results in the loss of the label, and thus a loss of signal.

Alternatively, as will be appreciated by those in the art, a labelled target sequence may be used as well; for example, a labelled primer may be used in the PCR amplification of the target, such that the label is incorporated in such a manner as to be cleaved off by the enzyme.

Alternatively, the readout probe (or, again, the target sequence) may comprise both a fluorescent label and a quencher, as is known in the art and depicted in Figure 6B. In this embodiment, the label and the quencher are attached to different nucleosides, yet are close enough that the quencher molecule results in little or no signal being present. Upon the introduction of the enzyme, the quencher is cleaved off, leaving the label, and allowing signalling by the label.

In addition, as will be appreciated by those in the art, these systems can be both solution-based assays or solid-phase assays, as outlined herein.

Furthermore, there are some systems that do not require cleavage for detection; for example, some nucleic acid binding proteins will bind to specific sequences and can thus serve as a secondary label. For example, some transcription factors will bind in a highly sequence dependent manner, and can distinguish between two SNPs. Having bound to the hybridization complex, a detectable binding partner can be added for detection.

In addition, as will be appreciated by those in the art, this type of approach works with other cleavage methods as well, for example the use of invasive cleavage methods, as outlined below.

Invasive cleavage

In a preferred embodiment, the determination of the identity of the base at the detection position of the target sequence proceeds using invasive cleavage technology. In general, invasive cleavage techniques rely on the use of structure-specific nucleases, where the structure can be formed as a result of the presence or absence of a mismatch. Generally, invasive cleavage technology may be described as follows. A target nucleic acid is recognized by two distinct probes. A first probe, generally referred to herein as an "invader" probe, is substantially complementary to a first portion of the target nucleic acid. A second probe, generally referred to herein as a "signal probe", is partially complementary to the target nucleic acid; the 3' end of the signal oligonucleotide is substantially complementary to the target sequence while the 5' end is non-complementary and preferably forms a

single-stranded "tail" or "arm". The non-complementary end of the second probe preferably comprises a "generic" or "unique" sequence, frequently referred to herein as a "detection sequence", that is used to indicate the presence or absence of the target nucleic acid, as described below. The detection sequence of the second probe preferably comprises at least one detectable label. Alternative methods have the detection sequence functioning as a target sequence for a capture probe, and thus rely on sandwich configurations using label probes.

Hybridization of the first and second oligonucleotides near or adjacent to one another on the target nucleic acid forms a number of structures. In a preferred embodiment, a forked cleavage structure, as shown in Figure 6, forms and is a substrate of a nuclease which cleaves the detection sequence from the signal oligonucleotide. The site of cleavage is controlled by the distance or overlap between the 3' end of the invader oligonucleotide and the downstream fork of the signal oligonucleotide. Therefore, neither oligonucleotide is subject to cleavage when misaligned or when unattached to target nucleic acid.

In a preferred embodiment, the nuclease that recognizes the forked cleavage structure and catalyzes release of the tail is thermostable, thereby, allowing thermal cycling of the cleavage reaction, if desired, although isothermal reactions are preferred. Preferred nucleases derived from thermostable DNA polymerases that have been modified to have reduced synthetic activity which is an unnecessary side-reaction during cleavage are disclosed in U.S. Patent Nos. 5,719,028 and 5,843,669, hereby expressly by reference. The synthetic activity of the DNA polymerase is reduced to a level where it does not interfere with detection of the cleavage reaction and detection of the freed tail. Preferably the DNA polymerase has no detectable polymerase activity. Examples of nucleases are those derived from *Thermus aquaticus*, *Thermus flavus*, or *Thermus thermophilus*.

In another embodiment, thermostable structure-specific nucleases are Flap endonucleases (FENs) selected from FEN-1 or FEN-2 like (e.g. XPG and RAD2 nucleases) from Archaeobacterial species, for example, FEN-1 from *Methanococcus jannaschii*, *Pyrococcus furiosus*, *Pyrococcus woesei*, and *Archaeoglobus fulgidus*. (U.S. Patent No. 5,843,669 and Lyamichev *et al.* 1999. Nature Biotechnology 17:292-297; both of which are hereby expressly by reference).

In a preferred embodiment, the nuclease is *AfuFEN1* or *PfuFEN1* nuclease. To cleave a forked structure, these nucleases require at least one overlapping nucleotide between the signal and invasive probes to recognize and cleave the 5' end of the signal probe. To effect cleavage the 3'-terminal nucleotide of the invader oligonucleotide is not required to be complementary to the target nucleic acid. In contrast, mismatch of the signal probe one base upstream of the cleavage site prevents creation of the overlap and cleavage. The specificity of the nuclease reaction allows single nucleotide

polymorphism (SNP) detection from, for example, genomic DNA, as outlined below (Lyamichev et al.).

The invasive cleavage assay is preferably performed on an array format. In a preferred embodiment, the signal probe has a detectable label, attached 5' from the site of nuclease cleavage (e.g. within the detection sequence) and a capture tag, as described herein for removal of the unreacted products (e.g. biotin or other hapten) 3' from the site of nuclease cleavage. After the assay is carried out, the uncleaved probe and the 3' portion of the cleaved signal probe (e.g. the detection sequence) may be extracted, for example, by binding to streptavidin beads or by crosslinking through the capture tag to produce aggregates or by antibody to an attached hapten. By "capture tag" herein is meant one of a pair of binding partners as described above, such as antigen/antibody pairs, digoxigenin, dinitrophenol, etc.

The cleaved 5' region, e.g. the detection sequence, of the signal probe, comprises a label and is detected and optionally quantitated. In one embodiment, the cleaved 5' region is hybridized to a probe on an array (capture probe) and optically detected (Figure 6). As described below, many different signal probes can be analyzed in parallel by hybridization to their complementary probes in an array. In a preferred embodiment as depicted in Figure 6, combination techniques are used to obtain higher specificity and reduce the detection of contaminating uncleaved signal probe or incorrectly cleaved product, an enzymatic recognition step is introduced in the array capture procedure. For example, as more fully outlined below, the cleaved signal probe binds to a capture probe to produce a double-stranded nucleic acid in the array. In this embodiment, the 3' end of the cleaved signal probe is adjacent to the 5' end of one strand of the capture probe, thereby, forming a substrate for DNA ligase (Broude et al. 1991. PNAS 91: 3072-3076). Only correctly cleaved product is ligated to the capture probe. Other incorrectly hybridized and non-cleaved signal probes are removed, for example, by heat denaturation, high stringency washes, and other methods that disrupt base pairing.

Accordingly, the present invention provides methods of determining the identity of a base at the detection position of a target sequence. In this embodiment, the target sequence comprises, 5' to 3', a first target domain comprising an overlap domain comprising at least a nucleotide in the detection position, and a second target domain contiguous with the detection position. A first probe (the "invader probe") is hybridized to the first target domain of the target sequence. A second probe (the "signal probe"), comprising a first portion that hybridizes to the second target domain of the target sequence and a second portion that does not hybridize to the target sequence, is hybridized to the second target domain. If the second probe comprises a base that is perfectly complementary to the detection position a cleavage structure is formed. The addition of a cleavage enzyme, such as is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,029; 5,541,311 and 5,843,669, all of which are expressly incorporated by reference, results in the cleavage of the detection sequence from the

signalling probe. This then can be used as a target sequence in an assay complex.

In addition, as for a variety of the techniques outlined herein, unreacted probes (i.e. signalling probes, in the case of invasive cleavage), may be removed using any number of techniques. For example, the use of a binding partner (70 in Figure 6C) coupled to a solid support comprising the other member of the binding pair can be done. Similarly, after cleavage of the primary signal probe, the newly created cleavage products can be selectively labeled at the 3' or 5' ends using enzymatic or chemical methods.

Again, as outlined above, the detection of the invasive cleavage reaction can occur directly, in the case where the detection sequence comprises at least one label, or indirectly, using sandwich assays, through the use of additional probes; that is, the detection sequences can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc.

In addition, as for most of the techniques outlined herein, these techniques may be done for the two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set for the other strand of the target.

Thus, the invasive cleavage reaction requires, in no particular order, an invader probe, a signalling probe, and a cleavage enzyme.

As for other methods outlined herein, the invasive cleavage reaction may be done as a solution based assay or a solid phase assay.

Solution-based invasive cleavage

The invasive cleavage reaction may be done in solution, followed by addition of one of the components to an array, with optional (but preferable) removal of unreacted probes. For example, as depicted in Figure 6C, the reaction is carried out in solution, using a capture tag (i.e. a member of a binding partner pair) that is separated from the label on the detection sequence with the cleavage site. After cleavage (dependent on the base at the detection position), the signalling probe is cleaved. The capture tag is used to remove the uncleaved probes (for example, using magnetic particles comprising the other member of the binding pair), and the remaining solution is added to the array. Figure 6C depicts the direct attachment of the detection sequence to the capture probe. In this embodiment, the detection sequence can effectively act as an adapter sequence. In alternate embodiments, as depicted in Figure 6D, the detection sequence is unlabelled and an additional label probe is used; as

outlined below, this can be ligated to the hybridization complex.

Solid-phase based assays

The invasive cleavage reaction can also be done as a solid-phase assay. As depicted in Figure 6A, the target sequence can be attached to the array using a capture probe (in addition, although not shown, the target sequence may be directly attached to the array). In a preferred embodiment, the signalling probe comprises both a fluorophore label (attached to the portion of the signalling probe that hybridizes to the target) and a quencher (generally on the detection sequence), with a cleavage site in between. Thus, in the absence of cleavage, very little signal is seen due to the quenching reaction. After cleavage, however, the detection sequence is removed, along with the quencher, leaving the unquenched fluorophore. Similarly, the invasive probe may be attached to the array, as depicted in Figure 6B.

In a preferred embodiment, the invasive cleavage reaction is configured to utilize a fluorophore-quencher reaction. A signalling probe comprising both a fluorophore and a quencher is attached to the bead. The fluorophore is contained on the portion of the signalling probe that hybridizes to the target sequence, and the quencher is contained on a portion of the signalling probe that is on the other side of the cleavage site (termed the "detection sequence" herein). In a preferred embodiment, it is the 3' end of the signalling probe that is attached to the bead (although as will be appreciated by those in the art, the system can be configured in a variety of different ways, including methods that would result in a loss of signal upon cleavage). Thus, the quencher molecule is located 5' to the cleavage site. Upon assembly of an assay complex, comprising the target sequence, an invader probe, and a signalling probe, and the introduction of the cleavage enzyme, the cleavage of the complex results in the disassociation of the quencher from the complex, resulting in an increase in fluorescence.

In this embodiment, suitable fluorophore-quencher pairs are as known in the art. For example, suitable quencher molecules comprise Dabcyl.

COMBINATION TECHNIQUES

It is also possible to combine two or more of these techniques to do genotyping, quantification, detection of sequences, etc.

Novel combination of competitive hybridization and extension

In a preferred embodiment, a combination of competitive hybridization and extension, particularly SBE, is used. This may be generally described as follows. In this embodiment, different extension primers comprising different bases at the readout position are used. These are hybridized to a target sequence under stringency conditions that favor perfect matches, and then an extension reaction is

done. Basically, the readout probe that has the match at the readout position will be preferentially extended for two reasons; first, the readout probe will hybridize more efficiently to the target (e.g. has a slower off rate), and the extension enzyme will preferentially add a nucleotide to a "hybridized" base. The reactions can then be detected in a number of ways, as outlined herein.

5

The system can take on a number of configurations, depending on the number of labels used, the use of adapters, whether a solution-based or surface-based assay is done, etc. Several preferred embodiments are shown in Figure 7.

10

In a preferred embodiment, at least two different readout probes are used, each with a different base at the readout position and each with a unique detectable label that allows the identification of the base at the readout position. As described herein, these detectable labels may be either primary or secondary labels, with primary labels being preferred. As for all the competitive hybridization reactions, a competition for hybridization exists with the reaction conditions being set to favor match over mismatch. When the correct match occurs, the 3' end of the hybridization complex is now double stranded and thus serves as a template for an extension enzyme to add at least one base to the probe, at a position adjacent to the readout position. As will be appreciated by those in the art, for most SNP analysis, the nucleotide next to the detection position will be the same in all the reactions.

15

20

In one embodiment, chain terminating nucleotides may be used; alternatively, non-terminating nucleotides may be used and multiple nucleotides may be added, if desired. The latter may be particularly preferred as an amplification step of sorts; if the nucleotides are labelled, the addition of multiple labels can result in signal amplification.

25

In a preferred embodiment, the nucleotides are analogs that allow separation of reacted and unreacted primers as described herein; for example, this may be done by using a nuclease blocking moiety to protect extended primers and allow preferential degradation of unextended primers or biotin (or iminobiotin) to preferentially remove the extended primers (this is done in a solution based assay, followed by elution and addition to the array).

30

As for the other reactions outlined herein, this may be done as a solution based assay, or a solid phase assay. Solution based assays are generally depicted in Figures 7A, 7B and 7C. In a solid phase reaction, an example of which is depicted in Figure 7D, the capture probe serves as the readout probe; in this embodiment, different positions on the array (e.g. different beads) comprise different readout probes. That is, at least two different capture/readout probes are used, with three and four also possible, depending on the allele. The reaction is run under conditions that favor the formation of perfect match hybridization complexes. In this embodiment, the dNTPs comprise a detectable label,

35

preferably a primary label such as a fluorophore. Since the competitive readout probes are spatially defined in the array, one fluorescent label can distinguish between the alleles; furthermore, it is the same nucleotide that is being added in the reaction, since it is the position adjacent to the SNP that is being extended. As for all the competitive assays, relative fluorescence intensity distinguishes between the alleles and between homozygosity and heterozygosity.

For both solution and solid phase reactions, adapters may be additionally used. In a preferred embodiment, as shown in Figure 7B for the solution based assay (although as will be appreciated by those in the art, a solid phase reaction may be done as well), adapters on the 5' ends of the readout probes are used, with identical adapters used for each allele. Each readout probe has a unique detectable label that allows the determination of the base at the readout position. After hybridization and extension, the readout probes are added to the array; the adapter sequences direct the probes to particular array locations, and the relative intensities of the two labels distinguishes between alleles.

Alternatively, as depicted in Figure 7C for the solution based assay (although as will be appreciated by those in the art, a solid phase reaction may be done as well), a different adapter may be used for each readout probe. In this embodiment, a single label may be used, since spatial resolution is used to distinguish the alleles by having a unique adapter attached to each allelic probe. After hybridization and extension, the readout probes are added to the array; the unique adapter sequences direct the probes to unique array locations. In this embodiment, it is the relative intensities of two array positions that distinguishes between alleles.

As will be appreciated by those in the art, any array may be used in this novel method, including both ordered and random arrays. In a preferred embodiment, the arrays may be made through spotting techniques, photolithographic techniques, printing techniques, or preferably are bead arrays.

Combination of competitive hybridization and invasive cleavage

In a preferred embodiment, a combination of competitive hybridization and invasive cleavage is done. As will be appreciated by those in the art, this technique is invasive cleavage as described above, with at least two sets of probes comprising different bases in the readout position. By running the reactions under conditions that favor hybridization complexes with perfect matches, different alleles may be distinguished.

In a preferred embodiment, this technique is done on bead arrays.

Novel combination of invasive cleavage and ligation

In a preferred embodiment, invasive cleavage and ligation is done, as is generally depicted in Figure 8.

In this embodiment, the specificity of the invasive cleavage reaction is used to detect the nucleotide in the detection position, and the specificity of the ligase reaction is used to ensure that only cleaved probes give a signal; that is, the ligation reaction is ?? confer an extra level of specificity.

5 The detection sequence, comprising a detectable label, of the signal probe is cleaved if the correct basepairing is present, as outlined above. The detection sequence then serves as the "target sequence" in a secondary reaction for detection; it is added to a capture probe on a microsphere. The capture probe in this case comprises a first double stranded portion and a second single stranded portion that will hybridize to the detection sequence. Again, preferred embodiments utilize adjacent
10 portions, although dNTPs and a polymerase to fill in the "gap" may also be done. A ligase is then added. As shown in Figure 8A, only if the signal probe has been cleaved will ligation occur. This may be detected as outlined herein; preferred embodiments utilize stringency conditions that will discriminate between the ligated and unligated systems.

15 As will be appreciated by those in the art, this system may take on a number of configurations, including solution based and solid based assays. In a preferred embodiment, as outlined above, the system is configured such that only if cleavage occurs will ligation happen. In a preferred embodiment, this may be done using blocking moieties; the technique can generally be described as follows. An invasive cleavage reaction is done, using a signalling probe that is blocked at the 3' end.
20 Following cleavage, which creates a free 3' terminus, a ligation reaction is done, generally using a template target and a second ligation probe comprising a detectable label. Since the signalling probe has a blocked 3' end, only those probes undergoing cleavage get ligated and labelled.

25 Alternatively, the orientations may be switched; in this embodiment, a free 5' phosphate is generated and is available for labeling.

Accordingly, in this embodiment, a solution invasive cleavage reaction is done (although as will be appreciated by those in the art, a support bound invasive cleavage reaction may be done as well).

30 As will be appreciated by those in the art, any array may be used in this novel method, including both ordered (predefined) and random arrays. In a preferred embodiment, the arrays may be made through spotting techniques, photolithographic techniques, printing techniques, or preferably are bead arrays.

35 Combination of invasive cleavage and extension

In a preferred embodiment, a combination of invasive cleavage and extension reactions are done, as generally depicted in Figure 9. The technique can generally be described as follows. An invasive cleavage reaction is done, using a signalling probe that is blocked at the 3' end. Following cleavage, which creates a free 3' terminus, an extension reaction is done (either enzymatically or chemically) to add a detectable label. Since the signalling probe has a blocked 3' end, only those probes undergoing cleavage get labelled.

Alternatively, the orientations may be switched, for example when chemical extension or labeling is done. In this embodiment, a free 5' phosphate is generated and is available for labeling.

In a preferred embodiment, the invasive cleavage reaction is configured as shown in Figure 9B. In this embodiment, the signalling probe is attached to the array at the 5' end (e.g. to the detection sequence) and comprises a blocking moiety at the 3' end. The blocking moiety serves to prevent any alteration (including either enzymatic alteration or chemical alteration) of the 3' end. Suitable blocking moieties include, but are not limited to, chain terminators, alkyl groups, halogens; basically any non-hydroxyl moiety.

Upon formation of the assay complex comprising the target sequence, the invader probe, and the signalling probe, and the introduction of the cleavage enzyme, the portion of the signalling probe comprising the blocking moiety is removed. As a result, a free 3' OH group is generated. This can be extended either enzymatically or chemically, to incorporate a detectable label. For example, enzymatic extension may occur. In a preferred embodiment, a non-templated extension occurs, for example, through the use of terminal transferase. Thus, for example, a modified dNTP may be incorporated, wherein the modification comprises the presence of a primary label such as a fluor, or a secondary label such as biotin, followed by the addition of a labeled streptavidin, for example. Similarly, the addition of a template (e.g. a secondary target sequence that will hybridize to the detection sequence attached to the bead) allows the use of any number of reactions as outlined herein, such as simple extension, SBE, pyrosequencing, OLA, etc. Again, this generally (but not always) utilizes the incorporation of a label into the growing strand.

Alternatively, as will be appreciated by those in the art, chemical labelling or extension methods may be used to label the 3' OH group.

As for all the combination methods, there are several advantages to this method. First of all, the absence of any label on the surface prior to cleavage allows a high signal-to-noise ratio. Additionally, the signalling probe need not contain any labels, thus making synthesis easier. Furthermore, because the target-specific portion of the signalling probe is removed during the assay, the remaining detection

sequence can be any sequence. This allows the use of a common sequence for all beads; even if different reactions are carried out in parallel on the array, the post-cleavage detection can be identical for all assays, thus requiring only one set of reagents. As will be appreciated by those in the art, it is also possible to have different detection sequences if required. In addition, since the label is attached post-cleavage, there is a great deal of flexibility in the type of label that may be incorporated. This can lead to significant signal amplification; for example, the use of highly labeled streptavidin bound to a biotin on the detection sequence can give an increased signal per detection sequence. Similarly, the use of enzyme labels such as alkaline phosphatase or horseradish peroxidase allow signal amplification as well.

A further advantage is the two-fold specificity that is built into the assay. By requiring specificity at the cleavage step, followed by specificity at the extension step, increased signal-to-noise ratios are seen.

As will be appreciated by those in the art, while generally described as a solid phase assay, this reaction may also be done in solution; this is similar to the solution-based SBE reactions, wherein the detection sequence serves as the extension primer. It should be noted that the arrays used to detect the invasive cleavage/extension reactions may be of any type, including, but not limited to, spotted and printed arrays, photolithographic arrays, and bead arrays.

Combination of ligation and extension

In a preferred embodiment, OLA and SBE are combined, as is sometimes referred to as "Genetic Bit" analysis and described in Nikforov et al., Nucleic Acid Res. 22:4167 (1994), hereby expressly incorporated by reference. In this embodiment, the two ligation probes do not hybridize adjacently; rather, they are separated by one or more bases. The addition of dNTPs and a polymerase, in addition to the ligation probes and the ligase, results in an extended, ligated probe. As for SBE, the dNTPs may carry different labels, or separate reactions can be run, if the SBE portion of the reaction is used for genotyping. Alternatively, if the ligation portion of the reaction is used for genotyping, either no extension occurs due to mismatch of the 3' base (such that the polymerase will not extend it), or no ligation occurs due to mismatch of the 5' base. As will be appreciated by those in the art, the reaction products are assayed using microsphere arrays. Again, as outlined herein, the assays may be solution based assays, with the ligated, extended probes being added to a microsphere array, or solid-phase assays. In addition, the unextended, unligated primers may be removed prior to detection as needed, as is outlined herein. Furthermore, adapter sequences may also be used as outlined herein for OLA.

Combination of competitive hybridization and ligation

In a preferred embodiment, a combination of competitive hybridization and ligation is done. As will be

appreciated by those in the art, this technique is OLA as described above, with at least two sets of probes comprising different bases in the readout position. By running the reactions under conditions that favor hybridization complexes with perfect matches, different alleles may be distinguished.

5 In one embodiment, LCR is used to genotype a single genomic locus by incorporating two sets of two optically labeled AS oligonucleotides and a detection oligonucleotide in the ligation reaction. The oligonucleotide ligation step discriminates between the AS oligonucleotides through the efficiency of ligation between an oligonucleotide with a correct match with the target nucleic acid versus a mismatch base in the target nucleic acid at the ligation site. Accordingly, a detection oligonucleotide
10 ligates efficiently to an AS oligonucleotide if there is complete base pairing at the ligation site. One 3' oligonucleotide (T base at 5' end) is optically labeled with FAM (green fluorescent dye) and the other 3' oligonucleotide (C base at 5' end) is labelled with TMR (yellow fluorescent dye). An A base in the target nucleic acid base pairs with the corresponding T resulting in efficient ligation of the FAM-labeled oligonucleotide. A G base in the target nucleic acid results in ligation of the TMR-labeled
15 oligonucleotide. TMR and FAM have distinct emission spectrums. Accordingly, the wavelength of the oligonucleotide ligated to the 5' detection oligonucleotide indicates the nucleotide and thus the genotype of the target nucleic acid.

In a preferred embodiment, this technique is done on bead arrays.

Combination of competitive hybridization and invasive cleavage

In a preferred embodiment, a combination of competitive hybridization and invasive cleavage is done. As will be appreciated by those in the art, this technique is invasive cleavage as described above, with at least two sets of probes (either the invader probes or the signalling probes) comprising different
25 bases in the readout position. By running the reactions under conditions that favor hybridization complexes with perfect matches, different alleles may be distinguished.

In a preferred embodiment, this technique is done on bead arrays.

ATTACHMENT OF TARGET SEQUENCES TO ARRAYS

As is generally described herein, there are a variety of methods that can be used to attach target sequences to the solid supports of the invention, particularly to the microspheres that are distributed on a surface of a substrate. Most of these methods generally rely on capture probes attached to the array. However, the attachment may be direct or indirect. Direct attachment includes those situations
35 wherein an endogenous portion of the target sequence hybridizes to the capture probe, or where the target sequence has been manipulated to contain exogenous adapter sequences that are added to the target sequence, for example during an amplification reaction. Alternatively, the target sequences

may be directly attached to the beads. Indirect attachment utilizes one or more secondary probes, termed a "capture extender probe". These methods are further described in "Addressing Arrays using Sequence Specific Adapters", filed October 22, 1999 (no U.S.S.N. received yet), herein incorporated by reference.

In a preferred embodiment, direct attachment is done, as is generally depicted in Figure 1A. In this embodiment, the target sequence comprises a first target domain that hybridizes to all or part of the capture probe.

In a preferred embodiment, direct attachment is accomplished through the use of adapter sequences. An "adapter sequence" as used herein is a sequence that is generally not native to the target sequence, i.e. is exogenous, but is added during an amplification reaction, such as PCR or any of the other amplification techniques; see "Addressing Arrays using Sequence Specific Adapters", filed October 22, 1999 (no U.S.S.N. received yet); PCT 97/31256 and EP 0 799 897 A1, all of which are expressly incorporated by reference. In this embodiment, one or more of the amplification primers comprises a first portion comprising the adapter sequence and a second portion comprising the primer sequence. Extending the amplification primer as is well known in the art results in target sequences that comprise the adapter sequences. The adapter sequences are designed to be substantially complementary to capture probes.

In a preferred embodiment, indirect attachment of the target sequence to the array is done, through the use of capture extender probes. "Capture extender" probes are generally depicted in Figure 1C, and other figures, and have a first portion that will hybridize to all or part of the capture probe, and a second portion that will hybridize to a first portion of the target sequence. Two capture extender probes may also be used. This has generally been done to stabilize assay complexes for example when the target sequence is large, or when large amplifier probes (particularly branched or dendrimer amplifier probes) are used.

When only capture probes are utilized, it is necessary to have unique capture probes for each target sequence; that is, the surface must be customized to contain unique capture probes; e.g. each bead comprises a different capture probe.

Alternatively, the use of adapter sequences and capture extender probes allow the creation of more "universal" surfaces. In a preferred embodiment, an array of different and usually artificial capture probes are made; that is, the capture probes do not have complementarity to known target sequences. The adapter sequences can then be added to any target sequences, or soluble capture extender probes are made; this allows the manufacture of only one kind of array, with the user able to

customize the array through the use of adapter sequences or capture extender probes. This then allows the generation of customized soluble probes, which as will be appreciated by those in the art is generally simpler and less costly.

5 As will be appreciated by those in the art, the length of the adapter sequences will vary, depending on the desired "strength" of binding and the number of different adapters desired. In a preferred embodiment, adapter sequences range from about 5 to about 500 basepairs in length, with from about 8 to about 100 being preferred, and from about 10 to about 50 being particularly preferred.

10 In one embodiment, microsphere arrays containing a single type of capture probe are made; in this embodiment, the capture extender probes are added to the beads prior to loading on the array. The capture extender probes may be additionally fixed or crosslinked, as necessary.

15 In a preferred embodiment, as outlined in Figure 1B, the capture probe comprises a component of assay, e.g. an invasive probe, a signalling probe, an extension primer, etc.; that is, after hybridization to the target sequence, it is the capture probe itself that is reacted upon during the reaction.

20 In one embodiment, capture probes are not used, and the target sequences are attached directly to the sites on the array. For example, libraries of clonal nucleic acids, including DNA and RNA, are used. In this embodiment, individual nucleic acids are prepared, generally using conventional methods (including, but not limited to, propagation in plasmid or phage vectors, amplification techniques including PCR, etc.). The nucleic acids are preferably arrayed in some format, such as a microtiter plate format, and either spotted or beads are added for attachment of the libraries.

25 Attachment of the clonal libraries (or any of the nucleic acids outlined herein) may be done in a variety of ways, as will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such as AminoLink or biotinylated nucleotides that can then be used to attach the nucleic acid to a surface, as well as affinity capture by hybridization), cross-linking, and electrostatic attachment, etc.

30 In a preferred embodiment, affinity capture is used to attach the clonal nucleic acids to the surface. For example, cloned nucleic acids can be derivatized, for example with one member of a binding pair, and the beads derivatized with the other member of a binding pair. Suitable binding pairs are as described herein for secondary labels and IBL/DBL pairs. For example, the cloned nucleic acids may be
35 biotinylated (for example using enzymatic incorporation of biotinylated nucleotides, for by photoactivated cross-linking of biotin). Biotinylated nucleic acids can then be captured on streptavidin-coated beads, as is known in the art. Similarly, other hapten-receptor combinations can be used, such

as digoxigenin and anti-digoxigenin antibodies. Alternatively, chemical groups can be added in the form of derivatized nucleotides, that can them be used to add the nucleic acid to the surface.

Preferred attachments are covalent, although even relatively weak interactions (i.e. non-covalent) can be sufficient to attach a nucleic acid to a surface, if there are multiple sites of attachment per each nucleic acid. Thus, for example, electrostatic interactions can be used for attachment, for example by having beads carrying the opposite charge to the bioactive agent.

Similarly, affinity capture utilizing hybridization can be used to attach cloned nucleic acids to beads. For example, as is known in the art, polyA+RNA is routinely captured by hybridization to oligo-dT beads; this may include oligo-dT capture followed by a cross-linking step, such as psoralen crosslinking). If the nucleic acids of interest do not contain a polyA tract, one can be attached by polymerization with terminal transferase, or via ligation of an oligoA linker, as is known in the art.

Alternatively, chemical crosslinking may be done, for example by photoactivated crosslinking of thymidine to reactive groups, as is known in the art.

In general, special methods are required to decode clonal arrays, as is more fully outlined below.

ASSAY

All of the above compositions and methods are directed to the determination of the identification of the base at one or more detection positions within a target nucleic acid. The detection systems of the present invention are based on the incorporation (or in some cases, of the deletion) of a detectable label into an assay complex on an array, based on the presence or absence of a mismatch.

Accordingly, the compositions and methods of the present invention are used to identify the nucleotide(s) at a detection position within the target sequence. As is outlined herein, this identification step can comprise a wide variety of techniques, including, but not limited to, straight hybridization techniques (including competitive hybridization and stringency control); extension techniques (SBE, sequencing by synthesis, allelic PCR); ligation techniques (OLA, LCR and SPOLA); cleavage techniques (invasive cleavage, endonuclease techniques); or combinations thereof.

Accordingly, the present invention provides methods and compositions useful in the detection of nucleic acids. As will be appreciated by those in the art, the compositions of the invention can take on a wide variety of configurations, as is generally outlined in the Figures. As is more fully outlined below, preferred systems of the invention work as follows. A target nucleic acid sequence is attached (via hybridization) to an array site. This attachment can be either directly to a capture probe on the

surface, or indirectly, using capture extender probes as outlined herein. In some embodiments, the target sequence itself comprises the labels. Alternatively, a label probe is then added, forming an assay complex. The attachment of the label probe may be direct (i.e. hybridization to a portion of the target sequence), or indirect (i.e. hybridization to an amplifier probe that hybridizes to the target sequence), with all the required nucleic acids forming an assay complex.

All of the methods and compositions herein are drawn to methods of determining the base at the detection position of a target nucleic acid, generally by having differential reactions occur depending on the presence or absence of a mismatch. The reaction products are generally detected on arrays, and particularly microsphere arrays, as is outlined herein.

Accordingly, the present invention provides array compositions comprising at least a first substrate with a surface comprising individual sites. By "array" or "biochip" herein is meant a plurality of nucleic acids in an array format; the size of the array will depend on the composition and end use of the array. Nucleic acids arrays are known in the art, and can be classified in a number of ways; both ordered arrays (e.g. the ability to resolve chemistries at discrete sites), and random arrays are included. Ordered arrays include, but are not limited to, those made using photolithography techniques (Affymetrix GeneChip™), spotting techniques (Synteni and others), printing techniques (Hewlett Packard and Rosetta), three dimensional "gel pad" arrays, etc. A preferred embodiment utilizes microspheres on a variety of substrates including fiber optic bundles, as are outlined in PCTs US98/21193, PCT US99/14387 and PCT US98/05025; WO98/50782; and U.S.S.N.s 09/287,573, 09/151,877, 09/256,943, 09/316,154, 60/119,323, 09/315,584; all of which are expressly incorporated by reference. While much of the discussion below is directed to the use of microsphere arrays on fiber optic bundles, any array format of nucleic acids on solid supports may be utilized.

Arrays containing from about 2 different bioactive agents (e.g. different beads, when beads are used) to many millions can be made, with very large arrays being possible. Generally, the array will comprise from two to as many as a billion or more, depending on the size of the beads and the substrate, as well as the end use of the array, thus very high density, high density, moderate density, low density and very low density arrays may be made. Preferred ranges for very high density arrays are from about 10,000,000 to about 2,000,000,000, with from about 100,000,000 to about 1,000,000,000 being preferred (all numbers being in square cm). High density arrays range about 100,000 to about 10,000,000, with from about 1,000,000 to about 5,000,000 being particularly preferred. Moderate density arrays range from about 10,000 to about 100,000 being particularly preferred, and from about 20,000 to about 50,000 being especially preferred. Low density arrays are generally less than 10,000, with from about 1,000 to about 5,000 being preferred. Very low density arrays are less than 1,000, with from about 10 to about 1000 being preferred, and from about 100 to

about 500 being particularly preferred. In some embodiments, the compositions of the invention may not be in array format; that is, for some embodiments, compositions comprising a single bioactive agent may be made as well. In addition, in some arrays, multiple substrates may be used, either of different or identical compositions. Thus for example, large arrays may comprise a plurality of smaller substrates.

In addition, one advantage of the present compositions is that particularly through the use of fiber optic technology, extremely high density arrays can be made. Thus for example, because beads of 200 μm or less (with beads of 200 nm possible) can be used, and very small fibers are known, it is possible to have as many as 40,000 or more (in some instances, 1 million) different elements (e.g. fibers and beads) in a 1 mm^2 fiber optic bundle, with densities of greater than 25,000,000 individual beads and fibers (again, in some instances as many as 50-100 million) per 0.5 cm^2 obtainable (4 million per square cm for 5 μ center-to-center and 100 million per square cm for 1 μ center-to-center).

By "substrate" or "solid support" or other grammatical equivalents herein is meant any material that can be modified to contain discrete individual sites appropriate for the attachment or association of beads and is amenable to at least one detection method. As will be appreciated by those in the art, the number of possible substrates is very large. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the substrates allow optical detection and do not themselves appreciably fluoresce.

Generally the substrate is flat (planar), although as will be appreciated by those in the art, other configurations of substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Preferred substrates include optical fiber bundles as discussed below, and flat planar substrates such as glass, polystyrene and other plastics and acrylics.

In a preferred embodiment, the substrate is an optical fiber bundle or array, as is generally described in U.S.S.N.s 08/944,850 and 08/519,062, PCT US98/05025, and PCT US98/09163, all of which are expressly incorporated herein by reference. Preferred embodiments utilize preformed unitary fiber optic arrays. By "preformed unitary fiber optic array" herein is meant an array of discrete individual

fiber optic strands that are co-axially disposed and joined along their lengths. The fiber strands are generally individually clad. However, one thing that distinguished a preformed unitary array from other fiber optic formats is that the fibers are not individually physically manipulatable; that is, one strand generally cannot be physically separated at any point along its length from another fiber strand.

At least one surface of the substrate is modified to contain discrete, individual sites for later association of microspheres. These sites may comprise physically altered sites, i.e. physical configurations such as wells or small depressions in the substrate that can retain the beads, such that a microsphere can rest in the well, or the use of other forces (magnetic or compressive), or chemically altered or active sites, such as chemically functionalized sites, electrostatically altered sites, hydrophobically/ hydrophilically functionalized sites, spots of adhesive, etc.

The sites may be a pattern, i.e. a regular design or configuration, or randomly distributed. A preferred embodiment utilizes a regular pattern of sites such that the sites may be addressed in the X-Y coordinate plane. "Pattern" in this sense includes a repeating unit cell, preferably one that allows a high density of beads on the substrate. However, it should be noted that these sites may not be discrete sites. That is, it is possible to use a uniform surface of adhesive or chemical functionalities, for example, that allows the attachment of beads at any position. That is, the surface of the substrate is modified to allow attachment of the microspheres at individual sites, whether or not those sites are contiguous or non-contiguous with other sites. Thus, the surface of the substrate may be modified such that discrete sites are formed that can only have a single associated bead, or alternatively, the surface of the substrate is modified and beads may go down anywhere, but they end up at discrete sites.

In a preferred embodiment, the surface of the substrate is modified to contain wells, i.e. depressions in the surface of the substrate. This may be done as is generally known in the art using a variety of techniques, including, but not limited to, photolithography, stamping techniques, molding techniques and microetching techniques. As will be appreciated by those in the art, the technique used will depend on the composition and shape of the substrate.

In a preferred embodiment, physical alterations are made in a surface of the substrate to produce the sites. In a preferred embodiment, the substrate is a fiber optic bundle and the surface of the substrate is a terminal end of the fiber bundle, as is generally described in 08/818,199 and 09/151,877, both of which are hereby expressly incorporated by reference. In this embodiment, wells are made in a terminal or distal end of a fiber optic bundle comprising individual fibers. In this embodiment, the cores of the individual fibers are etched, with respect to the cladding, such that small wells or depressions are formed at one end of the fibers. The required depth of the wells will depend on the size of the

beads to be added to the wells.

Generally in this embodiment, the microspheres are non-covalently associated in the wells, although the wells may additionally be chemically functionalized as is generally described below, cross-linking agents may be used, or a physical barrier may be used, i.e. a film or membrane over the beads.

In a preferred embodiment, the surface of the substrate is modified to contain chemically modified sites, that can be used to attach, either covalently or non-covalently, the microspheres of the invention to the discrete sites or locations on the substrate. "Chemically modified sites" in this context includes, but is not limited to, the addition of a pattern of chemical functional groups including amino groups, carboxy groups, oxo groups and thiol groups, that can be used to covalently attach microspheres, which generally also contain corresponding reactive functional groups; the addition of a pattern of adhesive that can be used to bind the microspheres (either by prior chemical functionalization for the addition of the adhesive or direct addition of the adhesive); the addition of a pattern of charged groups (similar to the chemical functionalities) for the electrostatic attachment of the microspheres, i.e. when the microspheres comprise charged groups opposite to the sites; the addition of a pattern of chemical functional groups that renders the sites differentially hydrophobic or hydrophilic, such that the addition of similarly hydrophobic or hydrophilic microspheres under suitable experimental conditions will result in association of the microspheres to the sites on the basis of hydroaffinity. For example, the use of hydrophobic sites with hydrophobic beads, in an aqueous system, drives the association of the beads preferentially onto the sites. As outlined above, "pattern" in this sense includes the use of a uniform treatment of the surface to allow attachment of the beads at discrete sites, as well as treatment of the surface resulting in discrete sites. As will be appreciated by those in the art, this may be accomplished in a variety of ways.

In a preferred embodiment, the compositions of the invention further comprise a population of microspheres. By "population" herein is meant a plurality of beads as outlined above for arrays. Within the population are separate subpopulations, which can be a single microsphere or multiple identical microspheres. That is, in some embodiments, as is more fully outlined below, the array may contain only a single bead for each capture probe; preferred embodiments utilize a plurality of beads of each type.

By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of capture probe and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphite, titanium dioxide,

latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and Teflon may all be used. *"Microsphere Detection Guide"* from Bangs Laboratories, Fishers IN is a helpful guide.

5 The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for either capture probe attachment or tag attachment. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

10 It should be noted that a key component of the invention is the use of a substrate/bead pairing that allows the association or attachment of the beads at discrete sites on the surface of the substrate, such that the beads do not move during the course of the assay.

15 Each microsphere comprises a capture probe, although as will be appreciated by those in the art, there may be some microspheres which do not contain a capture probe, depending on the synthetic methods.

20 Attachment of the nucleic acids may be done in a variety of ways, as will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such as AminoLink or biotinylated nucleotides that can then be used to attach the nucleic acid to a surface, as well as affinity capture by hybridization), cross-linking, and electrostatic attachment, etc. In a preferred embodiment, affinity capture is used to attach the nucleic acids to the beads. For example, nucleic acids can be derivatized, for example with one member of a binding pair, and the beads derivatized with the other member of a binding pair. Suitable binding pairs are as described herein for IBL/DBL pairs. For example, the nucleic acids may be biotinylated (for example using enzymatic incorporation of biotinylated nucleotides, for by photoactivated cross-linking of biotin). Biotinylated nucleic acids can then be captured on streptavidin-coated beads, as is known in the art. Similarly, other hapten-receptor combinations can be used, such as digoxigenin and anti-digoxigenin antibodies. Alternatively, chemical groups can be added in the form of derivatized nucleotides, that can then be used to add the nucleic acid to the surface.

30 Preferred attachments are covalent, although even relatively weak interactions (i.e. non-covalent) can be sufficient to attach a nucleic acid to a surface, if there are multiple sites of attachment per each nucleic acid. Thus, for example, electrostatic interactions can be used for attachment, for example by having beads carrying the opposite charge to the bioactive agent.

Similarly, affinity capture utilizing hybridization can be used to attach nucleic acids to beads. For example, as is known in the art, polyA+RNA is routinely captured by hybridization to oligo-dT beads; this may include oligo-dT capture followed by a cross-linking step, such as psoralen crosslinking). If the nucleic acids of interest do not contain a polyA tract, one can be attached by polymerization with terminal transferase, or via ligation of an oligoA linker, as is known in the art.

Alternatively, chemical crosslinking may be done, for example by photoactivated crosslinking of thymidine to reactive groups, as is known in the art.

In general, probes of the present invention are designed to be complementary to a target sequence (either the target sequence of the sample or to other probe sequences, as is described herein), such that hybridization of the target and the probes of the present invention occurs. This complementarity need not be perfect; there may be any number of base pair mismatches that will interfere with hybridization between the target sequence and the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under the selected reaction conditions.

In a preferred embodiment, each bead comprises a single type of capture probe, although a plurality of individual capture probes are preferably attached to each bead. Similarly, preferred embodiments utilize more than one microsphere containing a unique capture probe; that is, there is redundancy built into the system by the use of subpopulations of microspheres, each microsphere in the subpopulation containing the same capture probe.

As will be appreciated by those in the art, the capture probes may either be synthesized directly on the beads, or they may be made and then attached after synthesis. In a preferred embodiment, linkers are used to attach the capture probes to the beads, to allow both good attachment, sufficient flexibility to allow good interaction with the target molecule, and to avoid undesirable binding reactions.

In a preferred embodiment, the capture probes are synthesized directly on the beads. As is known in the art, many classes of chemical compounds are currently synthesized on solid supports, such as peptides, organic moieties, and nucleic acids. It is a relatively straightforward matter to adjust the current synthetic techniques to use beads.

In a preferred embodiment, the capture probes are synthesized first, and then covalently attached to the beads. As will be appreciated by those in the art, this will be done depending on the composition

of the capture probes and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the art. Accordingly, "blank" microspheres may be used that have surface chemistries that facilitate the attachment of the desired functionality by the user. Some examples of these surface chemistries for blank microspheres include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

When microsphere arrays are used, an encoding/decoding system must be used. That is, since the beads are generally put onto the substrate randomly, there are several ways to correlate the functionality on the bead with its location, including the incorporation of unique optical signatures, generally fluorescent dyes, that could be used to identify the chemical functionality on any particular bead. This allows the synthesis of the candidate agents (i.e. compounds such as nucleic acids and antibodies) to be divorced from their placement on an array, i.e. the candidate agents may be synthesized on the beads, and then the beads are randomly distributed on a patterned surface. Since the beads are first coded with an optical signature, this means that the array can later be "decoded", i.e. after the array is made, a correlation of the location of an individual site on the array with the bead or candidate agent at that particular site can be made. This means that the beads may be randomly distributed on the array, a fast and inexpensive process as compared to either the in situ synthesis or spotting techniques of the prior art.

However, the drawback to these methods is that for a large array, the system requires a large number of different optical signatures, which may be difficult or time-consuming to utilize. Accordingly, the present invention provides several improvements over these methods, generally directed to methods of coding and decoding the arrays. That is, as will be appreciated by those in the art, the placement of the capture probes is generally random, and thus a coding/decoding system is required to identify the probe at each location in the array. This may be done in a variety of ways, as is more fully outlined below, and generally includes: a) the use a decoding binding ligand (DBL), generally directly labeled, that binds to either the capture probe or to identifier binding ligands (IBLs) attached to the beads; b) positional decoding, for example by either targeting the placement of beads (for example by using photoactivatable or photocleavable moieties to allow the selective addition of beads to particular locations), or by using either sub-bundles or selective loading of the sites, as are more fully outlined below; c) selective decoding, wherein only those beads that bind to a target are decoded; or d) combinations of any of these. In some cases, as is more fully outlined below, this decoding may occur for all the beads, or only for those that bind a particular target sequence. Similarly, this may occur either prior to or after addition of a target sequence. In addition, as outlined herein, the target sequences detected may be either a primary target sequence (e.g. a patient sample), or a reaction

product from one of the methods described herein (e.g. an extended SBE probe, a ligated probe, a cleaved signal probe, etc.).

Once the identity (i.e. the actual agent) and location of each microsphere in the array has been fixed, the array is exposed to samples containing the target sequences, although as outlined below, this can be done prior to or during the analysis as well. The target sequences can hybridize (either directly or indirectly) to the capture probes as is more fully outlined below, and results in a change in the optical signal of a particular bead.

In the present invention, "decoding" does not rely on the use of optical signatures, but rather on the use of decoding binding ligands that are added during a decoding step. The decoding binding ligands will bind either to a distinct identifier binding ligand partner that is placed on the beads, or to the capture probe itself. The decoding binding ligands are either directly or indirectly labeled, and thus decoding occurs by detecting the presence of the label. By using pools of decoding binding ligands in a sequential fashion, it is possible to greatly minimize the number of required decoding steps.

In some embodiments, the microspheres may additionally comprise identifier binding ligands for use in certain decoding systems. By "identifier binding ligands" or "IBLs" herein is meant a compound that will specifically bind a corresponding decoder binding ligand (DBL) to facilitate the elucidation of the identity of the capture probe attached to the bead. That is, the IBL and the corresponding DBL form a binding partner pair. By "specifically bind" herein is meant that the IBL binds its DBL with specificity sufficient to differentiate between the corresponding DBL and other DBLs (that is, DBLs for other IBLs), or other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the decoding step, including wash steps to remove non-specific binding.

In some embodiments, for example when the IBLs and corresponding DBLs are proteins or nucleic acids, the dissociation constants of the IBL to its DBL will be less than about 10^{-4} - 10^{-6} M⁻¹, with less than about 10^{-5} to 10^{-9} M⁻¹ being preferred and less than about 10^{-7} - 10^{-9} M⁻¹ being particularly preferred.

IBL-DBL binding pairs are known or can be readily found using known techniques. For example, when the IBL is a protein, the DBLs include proteins (particularly including antibodies or fragments thereof (FABs, etc.)) or small molecules, or vice versa (the IBL is an antibody and the DBL is a protein). Metal ion- metal ion ligands or chelators pairs are also useful. Antigen-antibody pairs, enzymes and substrates or inhibitors, other protein-protein interacting pairs, receptor-ligands, complementary nucleic acids, and carbohydrates and their binding partners are also suitable binding pairs. Nucleic acid - nucleic acid binding proteins pairs are also useful. Similarly, as is generally described in U.S. Patents 5,270,163, 5,475,096, 5,567,588, 5,595,877, 5,637,459, 5,683,867, 5,705,337, and related

patents, hereby incorporated by reference, nucleic acid "aptamers" can be developed for binding to virtually any target; such an aptamer-target pair can be used as the IBL-DBL pair. Similarly, there is a wide body of literature relating to the development of binding pairs based on combinatorial chemistry methods.

In a preferred embodiment, the IBL is a molecule whose color or luminescence properties change in the presence of a selectively-binding DBL. For example, the IBL may be a fluorescent pH indicator whose emission intensity changes with pH. Similarly, the IBL may be a fluorescent ion indicator, whose emission properties change with ion concentration.

Alternatively, the IBL is a molecule whose color or luminescence properties change in the presence of various solvents. For example, the IBL may be a fluorescent molecule such as an ethidium salt whose fluorescence intensity increases in hydrophobic environments. Similarly, the IBL may be a derivative of fluorescein whose color changes between aqueous and nonpolar solvents.

In one embodiment, the DBL may be attached to a bead, i.e. a "decoder bead", that may carry a label such as a fluorophore.

In a preferred embodiment, the IBL-DBL pair comprise substantially complementary single-stranded nucleic acids. In this embodiment, the binding ligands can be referred to as "identifier probes" and "decoder probes". Generally, the identifier and decoder probes range from about 4 basepairs in length to about 1000, with from about 6 to about 100 being preferred, and from about 8 to about 40 being particularly preferred. What is important is that the probes are long enough to be specific, i.e. to distinguish between different IBL-DBL pairs, yet short enough to allow both a) dissociation, if necessary, under suitable experimental conditions, and b) efficient hybridization.

In a preferred embodiment, as is more fully outlined below, the IBLs do not bind to DBLs. Rather, the IBLs are used as identifier moieties ("IMs") that are identified directly, for example through the use of mass spectroscopy.

Alternatively, in a preferred embodiment, the IBL and the capture probe are the same moiety; thus, for example, as outlined herein, particularly when no optical signatures are used, the capture probe can serve as both the identifier and the agent. For example, in the case of nucleic acids, the bead-bound probe (which serves as the capture probe) can also bind decoder probes, to identify the sequence of the probe on the bead. Thus, in this embodiment, the DBLs bind to the capture probes.

In a preferred embodiment, the microspheres may contain an optical signature. That is, as outlined in

U.S.S.N.s 08/818,199 and 09/151,877, previous work had each subpopulation of microspheres comprising a unique optical signature or optical tag that is used to identify the unique capture probe of that subpopulation of microspheres; that is, decoding utilizes optical properties of the beads such that a bead comprising the unique optical signature may be distinguished from beads at other locations with different optical signatures. Thus the previous work assigned each capture probe a unique optical signature such that any microspheres comprising that capture probe are identifiable on the basis of the signature. These optical signatures comprised dyes, usually chromophores or fluorophores, that were entrapped or attached to the beads themselves. Diversity of optical signatures utilized different fluorochromes, different ratios of mixtures of fluorochromes, and different concentrations (intensities) of fluorochromes.

In a preferred embodiment, the present invention does not rely solely on the use of optical properties to decode the arrays. However, as will be appreciated by those in the art, it is possible in some embodiments to utilize optical signatures as an additional coding method, in conjunction with the present system. Thus, for example, as is more fully outlined below, the size of the array may be effectively increased while using a single set of decoding moieties in several ways, one of which is the use of optical signatures on some beads. Thus, for example, using one "set" of decoding molecules, the use of two populations of beads, one with an optical signature and one without, allows the effective doubling of the array size. The use of multiple optical signatures similarly increases the possible size of the array.

In a preferred embodiment, each subpopulation of beads comprises a plurality of different IBLs. By using a plurality of different IBLs to encode each capture probe, the number of possible unique codes is substantially increased. That is, by using one unique IBL per capture probe, the size of the array will be the number of unique IBLs (assuming no "reuse" occurs, as outlined below). However, by using a plurality of different IBLs per bead, n , the size of the array can be increased to 2^n , when the presence or absence of each IBL is used as the indicator. For example, the assignment of 10 IBLs per bead generates a 10 bit binary code, where each bit can be designated as "1" (IBL is present) or "0" (IBL is absent). A 10 bit binary code has 2^{10} possible variants. However, as is more fully discussed below, the size of the array may be further increased if another parameter is included such as concentration or intensity; thus for example, if two different concentrations of the IBL are used, then the array size increases as 3^n . Thus, in this embodiment, each individual capture probe in the array is assigned a combination of IBLs, which can be added to the beads prior to the addition of the capture probe, after, or during the synthesis of the capture probe, i.e. simultaneous addition of IBLs and capture probe components.

Alternatively, the combination of different IBLs can be used to elucidate the sequence of the nucleic

acid. Thus, for example, using two different IBLs (IBL1 and IBL2), the first position of a nucleic acid can be elucidated: for example, adenosine can be represented by the presence of both IBL1 and IBL2; thymidine can be represented by the presence of IBL1 but not IBL2, cytosine can be represented by the presence of IBL2 but not IBL1, and guanosine can be represented by the absence of both. The second position of the nucleic acid can be done in a similar manner using IBL3 and IBL4; thus, the presence of IBL1, IBL2, IBL3 and IBL4 gives a sequence of AA; IBL1, IBL2, and IBL3 shows the sequence AT; IBL1, IBL3 and IBL4 gives the sequence TA, etc. The third position utilizes IBL5 and IBL6, etc. In this way, the use of 20 different identifiers can yield a unique code for every possible 10-mer.

In this way, a sort of "bar code" for each sequence can be constructed; the presence or absence of each distinct IBL will allow the identification of each capture probe.

In addition, the use of different concentrations or densities of IBLs allows a "reuse" of sorts. If, for example, the bead comprising a first agent has a 1X concentration of IBL, and a second bead comprising a second agent has a 10X concentration of IBL, using saturating concentrations of the corresponding labelled DBL allows the user to distinguish between the two beads.

Once the microspheres comprising the capture probes are generated, they are added to the substrate to form an array. It should be noted that while most of the methods described herein add the beads to the substrate prior to the assay, the order of making, using and decoding the array can vary. For example, the array can be made, decoded, and then the assay done. Alternatively, the array can be made, used in an assay, and then decoded; this may find particular use when only a few beads need be decoded. Alternatively, the beads can be added to the assay mixture, i.e. the sample containing the target sequences, prior to the addition of the beads to the substrate; after addition and assay, the array may be decoded. This is particularly preferred when the sample comprising the beads is agitated or mixed; this can increase the amount of target sequence bound to the beads per unit time, and thus (in the case of nucleic acid assays) increase the hybridization kinetics. This may find particular use in cases where the concentration of target sequence in the sample is low; generally, for low concentrations, long binding times must be used.

In general, the methods of making the arrays and of decoding the arrays is done to maximize the number of different candidate agents that can be uniquely encoded. The compositions of the invention may be made in a variety of ways. In general, the arrays are made by adding a solution or slurry comprising the beads to a surface containing the sites for attachment of the beads. This may be done in a variety of buffers, including aqueous and organic solvents, and mixtures. The solvent can evaporate, and excess beads are removed.

In a preferred embodiment, when non-covalent methods are used to associate the beads with the array, a novel method of loading the beads onto the array is used. This method comprises exposing the array to a solution of particles (including microspheres and cells) and then applying energy, e.g. agitating or vibrating the mixture. This results in an array comprising more tightly associated particles, as the agitation is done with sufficient energy to cause weakly-associated beads to fall off (or out, in the case of wells). These sites are then available to bind a different bead. In this way, beads that exhibit a high affinity for the sites are selected. Arrays made in this way have two main advantages as compared to a more static loading: first of all, a higher percentage of the sites can be filled easily, and secondly, the arrays thus loaded show a substantial decrease in bead loss during assays. Thus, in a preferred embodiment, these methods are used to generate arrays that have at least about 50% of the sites filled, with at least about 75% being preferred, and at least about 90% being particularly preferred. Similarly, arrays generated in this manner preferably lose less than about 20% of the beads during an assay, with less than about 10% being preferred and less than about 5% being particularly preferred.

In this embodiment, the substrate comprising the surface with the discrete sites is immersed into a solution comprising the particles (beads, cells, etc.). The surface may comprise wells, as is described herein, or other types of sites on a patterned surface such that there is a differential affinity for the sites. This differential affinity results in a competitive process, such that particles that will associate more tightly are selected. Preferably, the entire surface to be "loaded" with beads is in fluid contact with the solution. This solution is generally a slurry ranging from about 10,000:1 beads:solution (vol:vol) to 1:1. Generally, the solution can comprise any number of reagents, including aqueous buffers, organic solvents, salts, other reagent components, etc. In addition, the solution preferably comprises an excess of beads; that is, there are more beads than sites on the array. Preferred embodiments utilize two-fold to billion-fold excess of beads.

The immersion can mimic the assay conditions; for example, if the array is to be "dipped" from above into a microtiter plate comprising samples, this configuration can be repeated for the loading, thus minimizing the beads that are likely to fall out due to gravity.

Once the surface has been immersed, the substrate, the solution, or both are subjected to a competitive process, whereby the particles with lower affinity can be disassociated from the substrate and replaced by particles exhibiting a higher affinity to the site. This competitive process is done by the introduction of energy, in the form of heat, sonication, stirring or mixing, vibrating or agitating the solution or substrate, or both.

A preferred embodiment utilizes agitation or vibration. In general, the amount of manipulation of the

substrate is minimized to prevent damage to the array; thus, preferred embodiments utilize the agitation of the solution rather than the array, although either will work. As will be appreciated by those in the art, this agitation can take on any number of forms, with a preferred embodiment utilizing microtiter plates comprising bead solutions being agitated using microtiter plate shakers.

5

The agitation proceeds for a period of time sufficient to load the array to a desired fill. Depending on the size and concentration of the beads and the size of the array, this time may range from about 1 second to days, with from about 1 minute to about 24 hours being preferred.

10

It should be noted that not all sites of an array may comprise a bead; that is, there may be some sites on the substrate surface which are empty. In addition, there may be some sites that contain more than one bead, although this is not preferred.

15

In some embodiments, for example when chemical attachment is done, it is possible to attach the beads in a non-random or ordered way. For example, using photoactivatable attachment linkers or photoactivatable adhesives or masks, selected sites on the array may be sequentially rendered suitable for attachment, such that defined populations of beads are laid down.

20

The arrays of the present invention are constructed such that information about the identity of the capture probe is built into the array, such that the random deposition of the beads in the fiber wells can be "decoded" to allow identification of the capture probe at all positions. This may be done in a variety of ways, and either before, during or after the use of the array to detect target molecules.

25

Thus, after the array is made, it is "decoded" in order to identify the location of one or more of the capture probes, i.e. each subpopulation of beads, on the substrate surface.

30

In a preferred embodiment, pyrosequencing techniques are used to decode the array, as is generally described in "Nucleic Acid Sequencing Using Microsphere Arrays", filed October 22, 1999 (no U.S.S.N. received yet), hereby expressly incorporated by reference.

35

In a preferred embodiment, a selective decoding system is used. In this case, only those microspheres exhibiting a change in the optical signal as a result of the binding of a target sequence are decoded. This is commonly done when the number of "hits", i.e. the number of sites to decode, is generally low. That is, the array is first scanned under experimental conditions in the absence of the target sequences. The sample containing the target sequences is added, and only those locations exhibiting a change in the optical signal are decoded. For example, the beads at either the positive or negative signal locations may be either selectively tagged or released from the array (for example

through the use of photocleavable linkers), and subsequently sorted or enriched in a fluorescence-activated cell sorter (FACS). That is, either all the negative beads are released, and then the positive beads are either released or analyzed in situ, or alternatively all the positives are released and analyzed. Alternatively, the labels may comprise halogenated aromatic compounds, and detection of the label is done using for example gas chromatography, chemical tags, isotopic tags mass spectral tags.

As will be appreciated by those in the art, this may also be done in systems where the array is not decoded; i.e. there need not ever be a correlation of bead composition with location. In this embodiment, the beads are loaded on the array, and the assay is run. The "positives", i.e. those beads displaying a change in the optical signal as is more fully outlined below, are then "marked" to distinguish or separate them from the "negative" beads. This can be done in several ways, preferably using fiber optic arrays. In a preferred embodiment, each bead contains a fluorescent dye. After the assay and the identification of the "positives" or "active beads", light is shown down either only the positive fibers or only the negative fibers, generally in the presence of a light-activated reagent (typically dissolved oxygen). In the former case, all the active beads are photobleached. Thus, upon non-selective release of all the beads with subsequent sorting, for example using a fluorescence activated cell sorter (FACS) machine, the non-fluorescent active beads can be sorted from the fluorescent negative beads. Alternatively, when light is shown down the negative fibers, all the negatives are non-fluorescent and the the positives are fluorescent, and sorting can proceed. The characterization of the attached capture probe may be done directly, for example using mass spectroscopy.

Alternatively, the identification may occur through the use of identifier moieties ("IMs"), which are similar to IBLs but need not necessarily bind to DBLs. That is, rather than elucidate the structure of the capture probe directly, the composition of the IMs may serve as the identifier. Thus, for example, a specific combination of IMs can serve to code the bead, and be used to identify the agent on the bead upon release from the bead followed by subsequent analysis, for example using a gas chromatograph or mass spectroscopy.

Alternatively, rather than having each bead contain a fluorescent dye, each bead comprises a non-fluorescent precursor to a fluorescent dye. For example, using photocleavable protecting groups, such as certain ortho-nitrobenzyl groups, on a fluorescent molecule, photoactivation of the fluorochrome can be done. After the assay, light is shown down again either the "positive" or the "negative" fibers, to distinguish these populations. The illuminated precursors are then chemically converted to a fluorescent dye. All the beads are then released from the array, with sorting, to form populations of fluorescent and non-fluorescent beads (either the positives and the negatives or vice

versa).

In an alternate preferred embodiment, the sites of attachment of the beads (for example the wells) include a photopolymerizable reagent, or the photopolymerizable agent is added to the assembled array. After the test assay is run, light is shown down again either the “positive” or the “negative” fibers, to distinguish these populations. As a result of the irradiation, either all the positives or all the negatives are polymerized and trapped or bound to the sites, while the other population of beads can be released from the array.

In a preferred embodiment, the location of every capture probe is determined using decoder binding ligands (DBLs). As outlined above, DBLs are binding ligands that will either bind to identifier binding ligands, if present, or to the capture probes themselves, preferably when the capture probe is a nucleic acid or protein.

In a preferred embodiment, as outlined above, the DBL binds to the IBL.

In a preferred embodiment, the capture probes are single-stranded nucleic acids and the DBL is a substantially complementary single-stranded nucleic acid that binds (hybridizes) to the capture probe, termed a decoder probe herein. A decoder probe that is substantially complementary to each candidate probe is made and used to decode the array. In this embodiment, the candidate probes and the decoder probes should be of sufficient length (and the decoding step run under suitable conditions) to allow specificity; i.e. each candidate probe binds to its corresponding decoder probe with sufficient specificity to allow the distinction of each candidate probe.

In a preferred embodiment, the DBLs are either directly or indirectly labeled. In a preferred embodiment, the DBL is directly labeled, that is, the DBL comprises a label. In an alternate embodiment, the DBL is indirectly labeled; that is, a labeling binding ligand (LBL) that will bind to the DBL is used. In this embodiment, the labeling binding ligand-DBL pair can be as described above for IBL-DBL pairs.

Accordingly, the identification of the location of the individual beads (or subpopulations of beads) is done using one or more decoding steps comprising a binding between the labeled DBL and either the IBL or the capture probe (i.e. a hybridization between the candidate probe and the decoder probe when the capture probe is a nucleic acid). After decoding, the DBLs can be removed and the array can be used; however, in some circumstances, for example when the DBL binds to an IBL and not to the capture probe, the removal of the DBL is not required (although it may be desirable in some circumstances). In addition, as outlined herein, decoding may be done either before the array is used

to in an assay, during the assay, or after the assay.

In one embodiment, a single decoding step is done. In this embodiment, each DBL is labeled with a unique label, such that the the number of unique tags is equal to or greater than the number of capture probes (although in some cases, "reuse" of the unique labels can be done, as described herein; similarly, minor variants of candidate probes can share the same decoder, if the variants are encoded in another dimension, i.e. in the bead size or label). For each capture probe or IBL, a DBL is made that will specifically bind to it and contains a unique tag, for example one or more fluorochromes. Thus, the identity of each DBL, both its composition (i.e. its sequence when it is a nucleic acid) and its label, is known. Then, by adding the DBLs to the array containing the capture probes under conditions which allow the formation of complexes (termed hybridization complexes when the components are nucleic acids) between the DBLs and either the capture probes or the IBLs, the location of each DBL can be elucidated. This allows the identification of the location of each capture probe; the random array has been decoded. The DBLs can then be removed, if necessary, and the target sample applied.

In a preferred embodiment, the number of unique labels is less than the number of unique capture probes, and thus a sequential series of decoding steps are used. In this embodiment, decoder probes are divided into n sets for decoding. The number of sets corresponds to the number of unique tags. Each decoder probe is labeled in n separate reactions with n distinct tags. All the decoder probes share the same n tags. The decoder probes are pooled so that each pool contains only one of the n tag versions of each decoder, and no two decoder probes have the same sequence of tags across all the pools. The number of pools required for this to be true is determined by the number of decoder probes and the n . Hybridization of each pool to the array generates a signal at every address. The sequential hybridization of each pool in turn will generate a unique, sequence-specific code for each candidate probe. This identifies the candidate probe at each address in the array. For example, if four tags are used, then $4 \times n$ sequential hybridizations can ideally distinguish 4^n sequences, although in some cases more steps may be required. After the hybridization of each pool, the hybrids are denatured and the decoder probes removed, so that the probes are rendered single-stranded for the next hybridization (although it is also possible to hybridize limiting amounts of target so that the available probe is not saturated. Sequential hybridizations can be carried out and analyzed by subtracting pre-existing signal from the previous hybridization).

An example is illustrative. Assuming an array of 16 probe nucleic acids (numbers 1-16), and four unique tags (four different fluors, for example; labels A-D). Decoder probes 1-16 are made that correspond to the probes on the beads. The first step is to label decoder probes 1-4 with tag A, decoder probes 5-8 with tag B, decoder probes 9-12 with tag C, and decoder probes 13-16 with tag D.

The probes are mixed and the pool is contacted with the array containing the beads with the attached candidate probes. The location of each tag (and thus each decoder and candidate probe pair) is then determined. The first set of decoder probes are then removed. A second set is added, but this time, decoder probes 1, 5, 9 and 13 are labeled with tag A, decoder probes 2, 6, 10 and 14 are labeled with tag B, decoder probes 3, 7, 11 and 15 are labeled with tag C, and decoder probes 4, 8, 12 and 16 are labeled with tag D. Thus, those beads that contained tag A in both decoding steps contain candidate probe 1; tag A in the first decoding step and tag B in the second decoding step contain candidate probe 2; tag A in the first decoding step and tag C in the second step contain candidate probe 3; etc. In one embodiment, the decoder probes are labeled in situ; that is, they need not be labeled prior to the decoding reaction. In this embodiment, the incoming decoder probe is shorter than the candidate probe, creating a 5' "overhang" on the decoding probe. The addition of labeled ddNTPs (each labeled with a unique tag) and a polymerase will allow the addition of the tags in a sequence specific manner, thus creating a sequence-specific pattern of signals. Similarly, other modifications can be done, including ligation, etc.

In addition, since the size of the array will be set by the number of unique decoding binding ligands, it is possible to "reuse" a set of unique DBLs to allow for a greater number of test sites. This may be done in several ways; for example, by using some subpopulations that comprise optical signatures. Similarly, the use of a positional coding scheme within an array; different sub-bundles may reuse the set of DBLs. Similarly, one embodiment utilizes bead size as a coding modality, thus allowing the reuse of the set of unique DBLs for each bead size. Alternatively, sequential partial loading of arrays with beads can also allow the reuse of DBLs. Furthermore, "code sharing" can occur as well.

In a preferred embodiment, the DBLs may be reused by having some subpopulations of beads comprise optical signatures. In a preferred embodiment, the optical signature is generally a mixture of reporter dyes, preferably fluorescent. By varying both the composition of the mixture (i.e. the ratio of one dye to another) and the concentration of the dye (leading to differences in signal intensity), matrices of unique optical signatures may be generated. This may be done by covalently attaching the dyes to the surface of the beads, or alternatively, by entrapping the dye within the bead.

In a preferred embodiment, the encoding can be accomplished in a ratio of at least two dyes, although more encoding dimensions may be added in the size of the beads, for example. In addition, the labels are distinguishable from one another; thus two different labels may comprise different molecules (i.e. two different fluors) or, alternatively, one label at two different concentrations or intensity.

In a preferred embodiment, the dyes are covalently attached to the surface of the beads. This may be done as is generally outlined for the attachment of the capture probes, using functional groups on the

surface of the beads. As will be appreciated by those in the art, these attachments are done to minimize the effect on the dye.

In a preferred embodiment, the dyes are non-covalently associated with the beads, generally by entrapping the dyes in the pores of the beads.

Additionally, encoding in the ratios of the two or more dyes, rather than single dye concentrations, is preferred since it provides insensitivity to the intensity of light used to interrogate the reporter dye's signature and detector sensitivity.

In a preferred embodiment, a spatial or positional coding system is done. In this embodiment, there are sub-bundles or subarrays (i.e. portions of the total array) that are utilized. By analogy with the telephone system, each subarray is an "area code", that can have the same tags (i.e. telephone numbers) of other subarrays, that are separated by virtue of the location of the subarray. Thus, for example, the same unique tags can be reused from bundle to bundle. Thus, the use of 50 unique tags in combination with 100 different subarrays can form an array of 5000 different capture probes. In this embodiment, it becomes important to be able to identify one bundle from another; in general, this is done either manually or through the use of marker beads, i.e. beads containing unique tags for each subarray.

In alternative embodiments, additional encoding parameters can be added, such as microsphere size. For example, the use of different size beads may also allow the reuse of sets of DBLs; that is, it is possible to use microspheres of different sizes to expand the encoding dimensions of the microspheres. Optical fiber arrays can be fabricated containing pixels with different fiber diameters or cross-sections; alternatively, two or more fiber optic bundles, each with different cross-sections of the individual fibers, can be added together to form a larger bundle; or, fiber optic bundles with fiber of the same size cross-sections can be used, but just with different sized beads. With different diameters, the largest wells can be filled with the largest microspheres and then moving onto progressively smaller microspheres in the smaller wells until all size wells are then filled. In this manner, the same dye ratio could be used to encode microspheres of different sizes thereby expanding the number of different oligonucleotide sequences or chemical functionalities present in the array. Although outlined for fiber optic substrates, this as well as the other methods outlined herein can be used with other substrates and with other attachment modalities as well.

In a preferred embodiment, the coding and decoding is accomplished by sequential loading of the microspheres into the array. As outlined above for spatial coding, in this embodiment, the optical signatures can be "reused". In this embodiment, the library of microspheres each comprising a

different capture probe (or the subpopulations each comprise a different capture probe), is divided into a plurality of sublibraries; for example, depending on the size of the desired array and the number of unique tags, 10 sublibraries each comprising roughly 10% of the total library may be made, with each sublibrary comprising roughly the same unique tags. Then, the first sublibrary is added to the fiber optic bundle comprising the wells, and the location of each capture probe is determined, generally through the use of DBLs. The second sublibrary is then added, and the location of each capture probe is again determined. The signal in this case will comprise the signal from the "first" DBL and the "second" DBL; by comparing the two matrices the location of each bead in each sublibrary can be determined. Similarly, adding the third, fourth, etc. sublibraries sequentially will allow the array to be filled.

In a preferred embodiment, codes can be "shared" in several ways. In a first embodiment, a single code (i.e. IBL/DBL pair) can be assigned to two or more agents if the target sequences differ sufficiently in their binding strengths. For example, two nucleic acid probes used in an mRNA quantitation assay can share the same code if the ranges of their hybridization signal intensities do not overlap. This can occur, for example, when one of the target sequences is always present at a much higher concentration than the other. Alternatively, the two target sequences might always be present at a similar concentration, but differ in hybridization efficiency.

Alternatively, a single code can be assigned to multiple agents if the agents are functionally equivalent. For example, if a set of oligonucleotide probes are designed with the common purpose of detecting the presence of a particular gene, then the probes are functionally equivalent, even though they may differ in sequence. Similarly, an array of this type could be used to detect homologs of known genes. In this embodiment, each gene is represented by a heterologous set of probes, hybridizing to different regions of the gene (and therefore differing in sequence). The set of probes share a common code. If a homolog is present, it might hybridize to some but not all of the probes. The level of homology might be indicated by the fraction of probes hybridizing, as well as the average hybridization intensity. Similarly, multiple antibodies to the same protein could all share the same code.

In a preferred embodiment, decoding of self-assembled random arrays is done on the bases of pH titration. In this embodiment, in addition to capture probes, the beads comprise optical signatures, wherein the optical signatures are generated by the use of pH-responsive dyes (sometimes referred to herein as "ph dyes") such as fluorophores. This embodiment is similar to that outlined in PCT US98/05025 and U.S.S.N. 09/151,877, both of which are expressly incorporated by reference, except that the dyes used in the present invention exhibit changes in fluorescence intensity (or other properties) when the solution pH is adjusted from below the pKa to above the pKa (or vice versa). In a preferred embodiment, a set of pH dyes are used, each with a different pKa, preferably separated by

at least 0.5 pH units. Preferred embodiments utilize a pH dye set of pKa's of 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11, and 11.5. Each bead can contain any subset of the pH dyes, and in this way a unique code for the capture probe is generated. Thus, the decoding of an array is achieved by titrating the array from pH 1 to pH 13, and measuring the fluorescence signal from each bead as a function of solution pH.

Thus, the present invention provides array compositions comprising a substrate with a surface comprising discrete sites. A population of microspheres is distributed on the sites, and the population comprises at least a first and a second subpopulation. Each subpopulation comprises a capture probe, and, in addition, at least one optical dye with a given pKa. The pKas of the different optical dyes are different.

In a preferred embodiment, "random" decoding probes can be made. By sequential hybridizations or the use of multiple labels, as is outlined above, a unique hybridization pattern can be generated for each sensor element. This allows all the beads representing a given clone to be identified as belonging to the same group. In general, this is done by using random or partially degenerate decoding probes, that bind in a sequence-dependent but not highly sequence-specific manner. The process can be repeated a number of times, each time using a different labeling entity, to generate a different pattern of signals based on quasi-specific interactions. In this way, a unique optical signature is eventually built up for each sensor element. By applying pattern recognition or clustering algorithms to the optical signatures, the beads can be grouped into sets that share the same signature (i.e. carry the same probes).

In order to identify the actual sequence of the clone itself, additional procedures are required; for example, direct sequencing can be done, or an ordered array containing the clones, such as a spotted cDNA array, to generate a "key" that links a hybridization pattern to a specific clone.

Alternatively, clone arrays can be decoded using binary decoding with vector tags. For example, partially randomized oligos are cloned into a nucleic acid vector (e.g. plasmid, phage, etc.). Each oligonucleotide sequence consists of a subset of a limited set of sequences. For example, if the limited set comprises 10 sequences, each oligonucleotide may have some subset (or all of the 10) sequences. Thus each of the 10 sequences can be present or absent in the oligonucleotide. Therefore, there are 2^{10} or 1,024 possible combinations. The sequences may overlap, and minor variants can also be represented (e.g. A, C, T and G substitutions) to increase the number of possible combinations. A nucleic acid library is cloned into a vector containing the random code sequences. Alternatively, other methods such as PCR can be used to add the tags. In this way it is possible to use a small number of oligo decoding probes to decode an array of clones.

As will be appreciated by those in the art, the systems of the invention may take on a large number of different configurations, as is generally depicted in the Figures. In general, there are three types of systems that can be used: (1) "non-sandwich" systems (also referred to herein as "direct" detection) in which the target sequence itself is labeled with detectable labels (again, either because the primers
5 comprise labels or due to the incorporation of labels into the newly synthesized strand); (2) systems in which label probes directly bind to the target analytes; and (3) systems in which label probes are indirectly bound to the target sequences, for example through the use of amplifier probes.

Detection of the reactions of the invention, including the direct detection of products and indirect
10 detection utilizing label probes (i.e. sandwich assays), is preferably done by detecting assay complexes comprising detectable labels, which can be attached to the assay complex in a variety of ways, as is more fully described below.

Once the target sequence has preferably been anchored to the array, an amplifier probe is hybridized
15 to the target sequence, either directly, or through the use of one or more label extender probes, which serves to allow "generic" amplifier probes to be made. As for all the steps outlined herein, this may be done simultaneously with capturing, or sequentially. Preferably, the amplifier probe contains a multiplicity of amplification sequences, although in some embodiments, as described below, the amplifier probe may contain only a single amplification sequence, or at least two amplification
20 sequences. The amplifier probe may take on a number of different forms; either a branched conformation, a dendrimer conformation, or a linear "string" of amplification sequences. Label probes comprising detectable labels (preferably but not required to be fluorophores) then hybridize to the amplification sequences (or in some cases the label probes hybridize directly to the target sequence), and the labels detected, as is more fully outlined below.

Accordingly, the present invention provides compositions comprising an amplifier probe. By "amplifier
25 probe" or "nucleic acid multimer" or "amplification multimer" or grammatical equivalents herein is meant a nucleic acid probe that is used to facilitate signal amplification. Amplifier probes comprise at least a first single-stranded nucleic acid probe sequence, as defined below, and at least one single-
30 stranded nucleic acid amplification sequence, with a multiplicity of amplification sequences being preferred.

Amplifier probes comprise a first probe sequence that is used, either directly or indirectly, to hybridize
35 to the target sequence. That is, the amplifier probe itself may have a first probe sequence that is substantially complementary to the target sequence, or it has a first probe sequence that is substantially complementary to a portion of an additional probe, in this case called a label extender probe, that has a first portion that is substantially complementary to the target sequence. In a

preferred embodiment, the first probe sequence of the amplifier probe is substantially complementary to the target sequence.

In general, as for all the probes herein, the first probe sequence is of a length sufficient to give specificity and stability. Thus generally, the probe sequences of the invention that are designed to hybridize to another nucleic acid (i.e. probe sequences, amplification sequences, portions or domains of larger probes) are at least about 5 nucleosides long, with at least about 10 being preferred and at least about 15 being especially preferred.

In a preferred embodiment, several different amplifier probes are used, each with first probe sequences that will hybridize to a different portion of the target sequence. That is, there is more than one level of amplification; the amplifier probe provides an amplification of signal due to a multiplicity of labelling events, and several different amplifier probes, each with this multiplicity of labels, for each target sequence is used. Thus, preferred embodiments utilize at least two different pools of amplifier probes, each pool having a different probe sequence for hybridization to different portions of the target sequence; the only real limitation on the number of different amplifier probes will be the length of the original target sequence. In addition, it is also possible that the different amplifier probes contain different amplification sequences, although this is generally not preferred.

In a preferred embodiment, the amplifier probe does not hybridize to the sample target sequence directly, but instead hybridizes to a first portion of a label extender probe. This is particularly useful to allow the use of "generic" amplifier probes, that is, amplifier probes that can be used with a variety of different targets. This may be desirable since several of the amplifier probes require special synthesis techniques. Thus, the addition of a relatively short probe as a label extender probe is preferred. Thus, the first probe sequence of the amplifier probe is substantially complementary to a first portion or domain of a first label extender single-stranded nucleic acid probe. The label extender probe also contains a second portion or domain that is substantially complementary to a portion of the target sequence. Both of these portions are preferably at least about 10 to about 50 nucleotides in length, with a range of about 15 to about 30 being preferred. The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target or probe sequences. For example, assuming a 5'-3' orientation of the complementary target sequence, the first portion may be located either 5' to the second portion, or 3' to the second portion. For convenience herein, the order of probe sequences are generally shown from left to right.

In a preferred embodiment, more than one label extender probe-amplifier probe pair may be used, that is, n is more than 1. That is, a plurality of label extender probes may be used, each with a portion that is substantially complementary to a different portion of the target sequence; this can serve as another

level of amplification. Thus, a preferred embodiment utilizes pools of at least two label extender probes, with the upper limit being set by the length of the target sequence.

5 In a preferred embodiment, more than one label extender probe is used with a single amplifier probe to reduce non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697, incorporated by reference herein. In this embodiment, a first portion of the first label extender probe hybridizes to a first portion of the target sequence, and the second portion of the first label extender probe hybridizes to a first probe sequence of the amplifier probe. A first portion of the second label extender probe hybridizes to a second portion of the target sequence, and the second portion of the second label
10 extender probe hybridizes to a second probe sequence of the amplifier probe. These form structures sometimes referred to as "cruciform" structures or configurations, and are generally done to confer stability when large branched or dendrimeric amplifier probes are used.

15 In addition, as will be appreciated by those in the art, the label extender probes may interact with a preamplifier probe, described below, rather than the amplifier probe directly.

Similarly, as outlined above, a preferred embodiment utilizes several different amplifier probes, each with first probe sequences that will hybridize to a different portion of the label extender probe. In addition, as outlined above, it is also possible that the different amplifier probes contain different
20 amplification sequences, although this is generally not preferred.

In addition to the first probe sequence, the amplifier probe also comprises at least one amplification sequence. An "amplification sequence" or "amplification segment" or grammatical equivalents herein is meant a sequence that is used, either directly or indirectly, to bind to a first portion of a label probe as is more fully described below. Preferably, the amplifier probe comprises a multiplicity of
25 amplification sequences, with from about 3 to about 1000 being preferred, from about 10 to about 100 being particularly preferred, and about 50 being especially preferred. In some cases, for example when linear amplifier probes are used, from 1 to about 20 is preferred with from about 5 to about 10 being particularly preferred.

30 The amplification sequences may be linked to each other in a variety of ways, as will be appreciated by those in the art. They may be covalently linked directly to each other, or to intervening sequences or chemical moieties, through nucleic acid linkages such as phosphodiester bonds, PNA bonds, etc., or through interposed linking agents such amino acid, carbohydrate or polyol bridges, or through other
35 cross-linking agents or binding partners. The site(s) of linkage may be at the ends of a segment, and/or at one or more internal nucleotides in the strand. In a preferred embodiment, the amplification sequences are attached via nucleic acid linkages.

In a preferred embodiment, branched amplifier probes are used, as are generally described in U.S. Patent No. 5,124,246, hereby incorporated by reference. Branched amplifier probes may take on "fork-like" or "comb-like" conformations. "Fork-like" branched amplifier probes generally have three or more oligonucleotide segments emanating from a point of origin to form a branched structure. The point of origin may be another nucleotide segment or a multifunctional molecule to which at least three segments can be covalently or tightly bound. "Comb-like" branched amplifier probes have a linear backbone with a multiplicity of sidechain oligonucleotides extending from the backbone. In either conformation, the pendant segments will normally depend from a modified nucleotide or other organic moiety having the appropriate functional groups for attachment of oligonucleotides. Furthermore, in either conformation, a large number of amplification sequences are available for binding, either directly or indirectly, to detection probes. In general, these structures are made as is known in the art, using modified multifunctional nucleotides, as is described in U.S. Patent Nos. 5,635,352 and 5,124,246, among others.

In a preferred embodiment, dendrimer amplifier probes are used, as are generally described in U.S. Patent No. 5,175,270, hereby expressly incorporated by reference. Dendrimeric amplifier probes have amplification sequences that are attached via hybridization, and thus have portions of double-stranded nucleic acid as a component of their structure. The outer surface of the dendrimer amplifier probe has a multiplicity of amplification sequences.

In a preferred embodiment, linear amplifier probes are used, that have individual amplification sequences linked end-to-end either directly or with short intervening sequences to form a polymer. As with the other amplifier configurations, there may be additional sequences or moieties between the amplification sequences. In one embodiment, the linear amplifier probe has a single amplification sequence.

In addition, the amplifier probe may be totally linear, totally branched, totally dendrimeric, or any combination thereof.

The amplification sequences of the amplifier probe are used, either directly or indirectly, to bind to a label probe to allow detection. In a preferred embodiment, the amplification sequences of the amplifier probe are substantially complementary to a first portion of a label probe. Alternatively, amplifier extender probes are used, that have a first portion that binds to the amplification sequence and a second portion that binds to the first portion of the label probe.

In addition, the compositions of the invention may include "preamplifier" molecules, which serves a bridging moiety between the label extender molecules and the amplifier probes. In this way, more

amplifier and thus more labels are ultimately bound to the detection probes. Preamplifier molecules may be either linear or branched, and typically contain in the range of about 30-3000 nucleotides.

Thus, label probes are either substantially complementary to an amplification sequence or to a portion of the target sequence.

Detection of the genotyping reactions of the invention, including the direct detection of genotyping products and indirect detection utilizing label probes (i.e. sandwich assays), is done by detecting assay complexes comprising labels.

In a preferred embodiment, several levels of redundancy are built into the arrays of the invention. Building redundancy into an array gives several significant advantages, including the ability to make quantitative estimates of confidence about the data and significant increases in sensitivity. Thus, preferred embodiments utilize array redundancy. As will be appreciated by those in the art, there are at least two types of redundancy that can be built into an array: the use of multiple identical sensor elements (termed herein "sensor redundancy"), and the use of multiple sensor elements directed to the same target analyte, but comprising different chemical functionalities (termed herein "target redundancy"). For example, for the detection of nucleic acids, sensor redundancy utilizes of a plurality of sensor elements such as beads comprising identical binding ligands such as probes. Target redundancy utilizes sensor elements with different probes to the same target: one probe may span the first 25 bases of the target, a second probe may span the second 25 bases of the target, etc. By building in either or both of these types of redundancy into an array, significant benefits are obtained. For example, a variety of statistical mathematical analyses may be done.

In addition, while this is generally described herein for bead arrays, as will be appreciated by those in the art, this techniques can be used for any type of arrays designed to detect target analytes. Furthermore, while these techniques are generally described for nucleic acid systems, these techniques are useful in the detection of other binding ligand/target analyte systems as well.

In a preferred embodiment, sensor redundancy is used. In this embodiment, a plurality of sensor elements, e.g. beads, comprising identical bioactive agents are used. That is, each subpopulation comprises a plurality of beads comprising identical bioactive agents (e.g. binding ligands). By using a number of identical sensor elements for a given array, the optical signal from each sensor element can be combined and any number of statistical analyses run, as outlined below. This can be done for a variety of reasons. For example, in time varying measurements, redundancy can significantly reduce the noise in the system. For non-time based measurements, redundancy can significantly increase the confidence of the data.

In a preferred embodiment, a plurality of identical sensor elements are used. As will be appreciated by those in the art, the number of identical sensor elements will vary with the application and use of the sensor array. In general, anywhere from 2 to thousands may be used, with from 2 to 100 being preferred, 2 to 50 being particularly preferred and from 5 to 20 being especially preferred. In general, preliminary results indicate that roughly 10 beads gives a sufficient advantage, although for some applications, more identical sensor elements can be used.

Once obtained, the optical response signals from a plurality of sensor beads within each bead subpopulation can be manipulated and analyzed in a wide variety of ways, including baseline adjustment, averaging, standard deviation analysis, distribution and cluster analysis, confidence interval analysis, mean testing, etc.

In a preferred embodiment, the first manipulation of the optical response signals is an optional baseline adjustment. In a typical procedure, the standardized optical responses are adjusted to start at a value of 0.0 by subtracting the integer 1.0 from all data points. Doing this allows the baseline-loop data to remain at zero even when summed together and the random response signal noise is canceled out. When the sample is a fluid, the fluid pulse-loop temporal region, however, frequently exhibits a characteristic change in response, either positive, negative or neutral, prior to the sample pulse and often requires a baseline adjustment to overcome noise associated with drift in the first few data points due to charge buildup in the CCD camera. If no drift is present, typically the baseline from the first data point for each bead sensor is subtracted from all the response data for the same bead. If drift is observed, the average baseline from the first ten data points for each bead sensor is subtracted from all the response data for the same bead. By applying this baseline adjustment, when multiple bead responses are added together they can be amplified while the baseline remains at zero. Since all beads respond at the same time to the sample (e.g. the sample pulse), they all see the pulse at the exact same time and there is no registering or adjusting needed for overlaying their responses. In addition, other types of baseline adjustment may be done, depending on the requirements and output of the system used.

Once the baseline has been adjusted, a number of possible statistical analyses may be run to generate known statistical parameters. Analyses based on redundancy are known and generally described in texts such as Freund and Walpole, Mathematical Statistics, Prentice Hall, Inc. New Jersey, 1980, hereby incorporated by reference in its entirety.

In a preferred embodiment, signal summing is done by simply adding the intensity values of all responses at each time point, generating a new temporal response comprised of the sum of all bead responses. These values can be baseline-adjusted or raw. As for all the analyses described herein,

signal summing can be performed in real time or during post-data acquisition data reduction and analysis. In one embodiment, signal summing is performed with a commercial spreadsheet program (Excel, Microsoft, Redmond, WA) after optical response data is collected.

5 In a preferred embodiment, cumulative response data is generated by simply adding all data points in successive time intervals. This final column, comprised of the sum of all data points at a particular time interval, may then be compared or plotted with the individual bead responses to determine the extent of signal enhancement or improved signal-to-noise ratios.

10 In a preferred embodiment, the mean of the subpopulation (i.e. the plurality of identical beads) is determined, using the well known Equation 1:

Equation 1

$$\mu = \sum \frac{x_i}{n}$$

In some embodiments, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

15

In a preferred embodiment, the standard deviation of the subpopulation can be determined, generally using Equation 2 (for the entire subpopulation) and Equation 3 (for less than the entire subpopulation):

Equation 2

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Equation 3

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

20

As for the mean, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

In a preferred embodiment, statistical analyses are done to evaluate whether a particular data point

has statistical validity within a subpopulation by using techniques including, but not limited to, t distribution and cluster analysis. This may be done to statistically discard outliers that may otherwise skew the result and increase the signal-to-noise ratio of any particular experiment. This may be done using Equation 4:

Equation 4

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

In a preferred embodiment, the quality of the data is evaluated using confidence intervals, as is known in the art. Confidence intervals can be used to facilitate more comprehensive data processing to measure the statistical validity of a result.

In a preferred embodiment, statistical parameters of a subpopulation of beads are used to do hypothesis testing. One application is tests concerning means, also called mean testing. In this application, statistical evaluation is done to determine whether two subpopulations are different. For example, one sample could be compared with another sample for each subpopulation within an array to determine if the variation is statistically significant.

In addition, mean testing can also be used to differentiate two different assays that share the same code. If the two assays give results that are statistically distinct from each other, then the subpopulations that share a common code can be distinguished from each other on the basis of the assay and the mean test, shown below in Equation 5:

Equation 5

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Furthermore, analyzing the distribution of individual members of a subpopulation of sensor elements may be done. For example, a subpopulation distribution can be evaluated to determine whether the distribution is binomial, Poisson, hypergeometric, etc.

In addition to the sensor redundancy, a preferred embodiment utilizes a plurality of sensor elements that are directed to a single target analyte but yet are not identical. For example, a single target

nucleic acid analyte may have two or more sensor elements each comprising a different probe. This adds a level of confidence as non-specific binding interactions can be statistically minimized. When nucleic acid target analytes are to be evaluated, the redundant nucleic acid probes may be overlapping, adjacent, or spatially separated. However, it is preferred that two probes do not compete for a single binding site, so adjacent or separated probes are preferred. Similarly, when proteinaceous target analytes are to be evaluated, preferred embodiments utilize bioactive agent binding agents that bind to different parts of the target. For example, when antibodies (or antibody fragments) are used as bioactive agents for the binding of target proteins, preferred embodiments utilize antibodies to different epitopes.

In this embodiment, a plurality of different sensor elements may be used, with from about 2 to about 20 being preferred, and from about 2 to about 10 being especially preferred, and from 2 to about 5 being particularly preferred, including 2, 3, 4 or 5. However, as above, more may also be used, depending on the application.

As above, any number of statistical analyses may be run on the data from target redundant sensors.

One benefit of the sensor element summing (referred to herein as "bead summing" when beads are used), is the increase in sensitivity that can occur.

Once made, the compositions of the invention find use in a number of applications. In a preferred embodiment, the compositions are used to probe a sample solution for the presence or absence of a target sequence, including the quantification of the amount of target sequence present.

For SNP analysis, the ratio of different labels at a particular location on the array indicates the homozygosity or heterozygosity of the target sample, assuming the same concentration of each readout probe is used. Thus, for example, assuming a first readout probe comprising a first base at the readout position with a first detectable label and a second readout probe comprising a second base at the readout position with a second detectable label, equal signals (roughly 1:1 (taking into account the different signal intensities of the different labels, different hybridization efficiencies, and other reasons)) of the first and second labels indicates a heterozygote. The absence of a signal from the first label (or a ratio of approximately 0:1) indicates a homozygote of the second detection base; the absence of a signal from the second label (or a ratio of approximately 1:0) indicates a homozygote for the first detection base. As is appreciated by those in the art, the actual ratios for any particular system are generally determined empirically.

The present invention also finds use as a methodology for the detection of mutations or mismatches in

target nucleic acid sequences. For example, recent focus has been on the analysis of the relationship between genetic variation and phenotype by making use of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of single nucleotide polymorphisms (SNPs), which occur at an average frequency of more than 1 per kilobase in human genomic DNA. Some SNPs, particularly those in and around coding sequences, are likely to be the direct cause of therapeutically relevant phenotypic variants. There are a number of well known polymorphisms that cause clinically important phenotypes; for example, the apoE2/3/4 variants are associated with different relative risk of Alzheimer's and other diseases (see Cordor et al., Science 261(1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to oligonucleotide arrays has been shown to be an accurate and reliable method of simultaneously genotyping at least hundreds of SNPs; see Wang et al., Science, 280:1077 (1998); see also Schafer et al., Nature Biotechnology 16:33-39 (1998). The compositions of the present invention may easily be substituted for the arrays of the prior art.

Generally, a sample containing a target analyte (whether for detection of the target analyte or screening for binding partners of the target analyte) is added to the array, under conditions suitable for binding of the target analyte to at least one of the capture probes, i.e. generally physiological conditions. The presence or absence of the target analyte is then detected. As will be appreciated by those in the art, this may be done in a variety of ways, generally through the use of a change in an optical signal. This change can occur via many different mechanisms. A few examples include the binding of a dye-tagged analyte to the bead, the production of a dye species on or near the beads, the destruction of an existing dye species, a change in the optical signature upon analyte interaction with dye on bead, or any other optical interrogatable event.

In a preferred embodiment, the change in optical signal occurs as a result of the binding of a target analyte that is labeled, either directly or indirectly, with a detectable label, preferably an optical label such as a fluorochrome. Thus, for example, when a proteinaceous target analyte is used, it may be either directly labeled with a fluor, or indirectly, for example through the use of a labeled antibody. Similarly, nucleic acids are easily labeled with fluorochromes, for example during PCR amplification as is known in the art. Alternatively, upon binding of the target sequences, a hybridization indicator may be used as the label. Hybridization indicators preferentially associate with double stranded nucleic acid, usually reversibly. Hybridization indicators include intercalators and minor and/or major groove binding moieties. In a preferred embodiment, intercalators may be used; since intercalation generally only occurs in the presence of double stranded nucleic acid, only in the presence of target hybridization will the label light up. Thus, upon binding of the target analyte to a capture probe, there is a new optical signal generated at that site, which then may be detected.

Alternatively, in some cases, as discussed above, the target analyte such as an enzyme generates a species that is either directly or indirectly optical detectable.

Furthermore, in some embodiments, a change in the optical signature may be the basis of the optical signal. For example, the interaction of some chemical target analytes with some fluorescent dyes on the beads may alter the optical signature, thus generating a different optical signal.

As will be appreciated by those in the art, in some embodiments, the presence or absence of the target analyte may be done using changes in other optical or non-optical signals, including, but not limited to, surface enhanced Raman spectroscopy, surface plasmon resonance, radioactivity, etc.

The assays may be run under a variety of experimental conditions, as will be appreciated by those in the art. A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding. Various blocking and washing steps may be utilized as is known in the art.

In addition, the present invention provides kits for the reactions of the invention, comprising components of the assays as outlined herein. In addition, a variety of other reagents may be included in the assays or the kits. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite activity.

All references cited herein are incorporated by reference in their entirety.